Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

All-Or-None and Conservation Effects in the Learning and Retention of Paired Associates: W. K. ESTES, B. L. HOPKINS, AND E. J. CROTHERS 329

Effects of Task Definition and Probability of Reinforcement upon the Acquisition and Extinction of Imitative Responses: V. T. KANAREFF AND J. T. LANZETTA 340

Acquired Drive Strength as a Joint Function of Shock Intensity and Number of Acquisition Trials: M. L. GOLDSTEIN 349

Instrumental Licking Behavior as a Function of Schedule, Volume, and Concentration of a Saccharine Reinforcer: S. H. HULSE, H. L. SNYDER, AND W. E. BACON 359

Drive, Verbal Performance, and Muscle Action Potential: J. B. SIDOWSKI AND R. G. EASON 365

Effects of Intermittent Reinforcement of an Irrelevant Dimension and Task Complexity upon Concept Identification: L. E. BOURNE, JR. AND R. C. HAYGOOD 371

Verbal Satiation and Changes in the Intensity of Meaning: W. E. LAMBERT AND L. A. JAKOBOVITS 376

Deprivation and Reward Magnitude Effects on Speed throughout the Goal Gradient: R. F. WEISS 384

Some Determinants of Two-Dimensional Visual Tracking Behavior: J. A. ADAMS AND L. V. XHIGNESSE 391

Individual Differences in Learning as a Function of Shock Level: R. KAPLAN, S. KAPLAN, AND E. L. WALKER 404

Comparison of Verbal Response Transfer Mediated by Meaningfully Similar and Associated Stimuli: J. J. RYAN 408

Supplementary Report: Alteration in the Reinforcement Value of a Positive Reinforcer: A. M. BUCHWALD 416

Replication Report: Verbal Context and the Recall of Meaningful Material: P. RICHARDSON AND J. F. VOSS 417

American Psychological Association

Vol. 60, No. 6 December 1960

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1960 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor | HELEN ORR, Promotion Manager | BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1960 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 60, NO. 6 DECEMBER 1960

ALL-OR-NONE AND CONSERVATION EFFECTS IN THE LEARNING AND RETENTION OF PAIRED ASSOCIATES[1]

W. K. ESTES, B. L. HOPKINS, AND E. J. CROTHERS

Indiana University

Tacitly in the functionalist approach (McGeoch & Irion, 1952) and explicitly in the more formalized theories (Estes, 1959a; Gibson, 1940; Hull, 1943; Hull, Hovland, Ross, Hall, Perkins, & Fitch, 1940; Spence, 1955) it has been assumed that simple associative learning proceeds by a gradual increase in the strengths of stimulus-response associations over a series of reinforced trials. This "incremental" assumption, with its implication that the probability of a correct response on the part of an individual *S* should be increased by each reinforcement, seems to be incorporated almost routinely into each new learning model that appears (e.g., Bourne & Restle, 1959; Bush & Mosteller, 1955; Luce, 1959).

As might be expected in the case of such a widely held conception, the empirical basis of the incremental assumption does not lack breadth. Nevertheless, it may not be quite solid, for at both the theoretical and experimental levels, small fissures have appeared. Some refractory problems having to do with rates and asymptotes of discrimination learning (Estes, 1959a) led the writer a number of years ago to consider a modified statistical model for learning in which the total pattern of stimulation occurring on a given trial is assumed to become associated with the reinforced response on an all-or-none basis (Estes, 1959b). The purpose of the present study is to provide a differential test of the "pattern model" against the class of models predicated on the incremental assumption.

Working in a different context, Rock (1957) has recently brought forward data which appear to challenge the incremental assumption, at least for paired-associate learning. Rock used a novel procedure in which incorrect items were dropped from the list and replaced with new ones following each presentation cycle. If the missed items were "partially learned," then *S*s learning under this

[1] The first author is responsible for the planning and interpretation of the study and for the present report. The second author collaborated in Exp. I and the third author, whose participation was made possible by a grant from the Ford Foundation, collaborated in Exp. II.

condition would be handicapped in comparison with conventional controls, but Rock found no difference in rate of learning under the two conditions. However, the interpretation of Rock's findings is complicated by the fact that the items eliminated under his procedure are probably more difficult to learn on the average than the items that replace them. In the hope of clarifying this somewhat perplexing situation, Exp. I of the present study was designed to attack the same problem with a method which would permit simpler controlled comparisons relative to the effects of single reinforced trials upon individual *S*s. Once this had been accomplished, it became apparent that the technique contrived to investigate possible all-or-none effects in acquisition could be extended to do the same for retention; Exp. II of this study is addressed to the latter problem.

Experiment I

The observed result of a single paired-associate trial is to increase the relative frequency of correct responses. With the 8-item list used in this study, the initial probability of a correct response to any item is at most .125. Under the conditions used, the effect of a single reinforcement (paired presentation of stimulus and response members of an item[2]) is to raise the relative frequency of correct responses to approximately .5 for a group of *S*s. The first question to be raised is which of the following interpretations is more nearly correct with reference to any given item: (*a*) the probability of a correct response is now .5 for each *S*, or (*b*) the probability of a correct response is now unity for half the *S*s and zero (or at least no greater than chance) for the remainder. To distinguish between these alternatives, a single reinforced trial will be followed by a test and then, without intervening reinforcement, a second test. The design may be represented as

R T T

where R denotes a single exposure of each stimulus-response pair in the list and T an exposure of each stimulus alone. If the first interpretation is correct, then correct and incorrect responses on the first test should be followed about equally often by correct responses on the second test. If the second interpretation is correct, the relative frequency of correct responses on the second test should be greater following correct responses on the first test than following incorrect responses on the first test. In the latter case, if there were no retention loss, the probability of correct following correct should be unity; forgetting would reduce this value by an amount that cannot be specified in advance. The probability of correct following incorrect is critical: according to (*a*) it must be greater than chance whereas according to (*b*) it must not.

As a control measure, the design will be completed by giving each *S* additional items with the paradigm

R T R T

permitting evaluation of the effect of a second reinforcement upon items

[2] To avoid confusion in the sequel, it should be emphasized that this definition of "reinforcement" is different from the definition in terms of "correct anticipation" that has been used, e.g., by Underwood (1954). The latter is inapplicable under the conditions of the present study since the test with the stimulus member of an item does not occur in temporal contiguity with the paired stimulus-response presentation, and learning, as measured in terms of frequency of correct responding, increases as a function of paired presentations, not as a function of preceding test trials.

which do not yield correct responses on the first test.

Method

Subjects and apparatus.—The *S*s were 48 students from introductory psychology classes at Indiana University who served as part of a course requirement.

From *S*'s view the apparatus appeared as an 18 × 18-in. gray panel containing two windows. The windows, each 2½ in. wide by 1½ in. high, were located 8¼ in. from the top of the panel and centered 1 in. apart horizontally. Each window was fitted with semisilvered glass, and behind each was a light-tight compartment containing two 110-v., 6-w. lamps which were turned on and off by means of a cam timer. The room was illuminated by a 60-w. lamp so placed above and behind *S*'s chair that *S* could not see through the windows when the compartment lights were off but could read typewritten material located 1 in. behind a window when its compartment lights were on.

The materials presented for paired-associate learning were two lists of eight consonant syllable-number pairs. The responses were the numbers 1 through 8 for each list. The syllables were randomly drawn sets of three consonants. Each pair was typed in elite type on a 5 × 8-in. white card and spaced so that when the card was in position behind the panel the syllable was centered behind the window on *S*'s left and the number behind the window on *S*'s right.

Procedure.—At the beginning of a session, *S* was seated before the exposure panel and given instructions to the effect that (*a*) on each practice trial both members of a pair would be shown; (*b*) *S*'s task was to learn to associate the appropriate number with each set of letters; (*c*) every so often, there would be test trials on which only the letters were shown, and *S* was to give the numbers from memory if possible; (*d*) on test trials *S* was to guess if not sure of the correct number.

For each list, *S* was given two practice trials and two test trials. On the first practice trial, all eight pairs were presented in a random order. Then the first test trial was given, all eight syllables being presented in a new random order. On the second practice trial, only four pairs were shown. Then on the second test trial, all eight syllables were presented in a new random order. The four pairs presented on the second practice trial were randomly selected for pairs of *S*s, one *S* of the pair being presented with the selected four pairs and the other *S* being presented with the four pairs withheld from his "partner." The same procedure was used for both lists, the order of presentation of the lists being counterbalanced over *S*s.

On practice trials, both windows were illuminated for 4 sec., then neither was illuminated during the 4-sec. interpair interval. On test trials, the left window, exposing the syllable alone, was illuminated for 4 sec., then neither window was illuminated during the 4-sec. interpair interval. There was a pause of 1 min. between lists.

TABLE 1

FREQUENCIES OF CORRECT (C) AND INCORRECT (N) RESPONSES ON TEST TRIALS 1 AND 2

One Reinforcement		Two Reinforcements	
T_1	T_2	T_1	T_2
C 188	C 133 N 55	C 155	C 134 N 21
N 196	C 18 N 178	N 229	C 106 N 123

TABLE 2

OBSERVED AND THEORETICAL (PATTERN MODEL) PROPORTIONS OF C ON TEST 2 GIVEN C OR N ON TEST 1

	One Reinforcement		Two Reinforcements	
	Observed	Theory	Observed	Theory
$C_2:C_1$	.71	.87	.86	.85
$C_2:N_1$	.09	.12	.46	.40

Results

Response frequencies pooled over *S*s and lists for each test trial are presented in Table 1. The data for the items having only one reinforcement are of principal interest. Nearly 50% of these were correct on the first test and nearly 40% on the second test. Proportions of cases in which items correct (C) or noncorrect, including omissions, (N) on the first test had correct responses on the second test

are shown in Table 2. The notation $C_2:C_1$ is to be read "correct on Test 2 given that the item was correct on Test 1," and $C_2:N_1$ "correct on Test 2 given that the item was not correct on Test 1." The observed difference between these two proportions is so large as to make wholly untenable the hypothesis that probabilities of correct responses were raised to approximately .5 by the single reinforcement for all *Ss* and items. Still more difficult to reconcile with an incremental assumption is the observation that the proportion for $C_2:N_1$ is even smaller than the .125 that could be achieved if *Ss* guessed at all unlearned items, using all of the response numbers with equal probabilities.

The simplest interpretation of this result would seem to be that in the case of items not correct on Test 1 the reinforcement had had no strengthening effect whatsoever. Several possible objections to this interpretation should be considered. Since *Ss* may not always have guessed when they did not know the correct answers, the guessing allowance of .125 may be too high. Also, it should be noted that if *Ss* had a strong tendency to repeat particular errors, the result would be a reduction in the proportion of C following N responses. These points can be taken care of by a new analysis using only items having incorrect responses (excluding omissions) on Test 1 and different responses (again excluding omissions) on Test 2. For these items, a proportion of .14 correct responses would be the chance expectation on Test 2. If it were the case either that the items incorrect on Test 1 were partially learned or that *Ss* restricted their guesses on Test 2 to the response members of unlearned items, then the proportion correct would be greater than .14. The observed proportion, based on 121 observations, was .13.

Other objections might be based on the possibility that the instances of C responses on Test 1 represent either markedly easier items or *Ss* of much greater learning ability than instances of N responses on Test 1. The possibility of differential learning rates can be investigated quite directly. The instances of $C_2:N_1$ were scattered over 17 different *Ss*. Computing the proportion of C responses on Test 1 and also the proportion $C_2:C_1$ for these *Ss* yields values of .44 and .70, respectively. Comparing these with the corresponding values of .49 and .71 obtained for the full group of 48 *Ss*, one must conclude that individual differences could have played no significant role in generating the row differences appearing in Table 2. If differences in item difficulty were an important factor, then for the twice reinforced items having N responses on Test 1 we should find that the second reinforcement, given prior to Test 2, would produce only a small increment in probability of correct response as compared to the effect of the first reinforcement upon this probability in the full population of items. However, Table 2 shows the proportion $C_2:N_1$ to be .46 for the twice-reinforced items. Since the same set of items that yielded this proportion of .46 correct after the second reinforcement had manifested no greater than a chance probability of correct response after the first reinforcement, there appears to be no support here for a hypothesis involving important differences in difficulty between the items having C and those having N responses on Test 1.

These analyses having revealed no evidence of any artifact that could account for the observed pattern of response proportions, we are left with no prepossessing alternative to an interpretation in terms of all-or-none acquisition.

In one respect, however, our data do not fit the conventional picture of all-or-none learning; for, once a correct response has occurred to a given item, it does not inevitably recur on the next test trial. This observation of less than perfect retention may mean that an association learned on a reinforced trial does not necessarily stay learned over a subsequent series of rest intervals and

unreinforced tests. But before this conclusion can be justified, the contribution of guessing shrinkage must be evaluated. If on both tests, *Ss* guessed on unlearned items, then apparent forgetting would occur in cases when responses to particular items were correct by chance on the first test but not on the second. To provide a basis for appraising this factor, the theoretical values shown in Table 2 have been calculated from the "pattern model" mentioned above. The pertinent assumptions are simply that on each reinforced trial each unlearned item has some fixed probability c of being learned and that learned items always yield correct responses on subsequent tests. With the boundary condition that unlearned items have probability $\frac{1}{8}$ of yielding correct responses on test trials, for an 8-item list, we can estimate the parameter c from the observed proportions of correct responses on the first test. For the singly reinforced items, for example, this proportion was .49, from which we obtain

$$.49 = c + (1 - c)\tfrac{1}{8}$$
$$c = \frac{.365}{.875} = .42$$

A similar determination for the twice reinforced items yields a value of .31 for c, the difference between the two estimates being presumably attributable to sampling error.

Using these estimates of the conditioning probability, the values shown in Table 2 were calculated for the conditional probabilities of $C_2{:}C_1$ and $C_2{:}N_1$ for the once- and twice-reinforced items. It appears that both theoretical values are quite satisfactory for the twice reinforced items, suggesting that the model represents the learning process quite well when further reinforcements following the first test are available to offset any retention losses. For the once reinforced items, the predicted probability of $C_2{:}N_1$ is close to the mark but that of $C_2{:}C_1$ is considerably too high. It appears that there must be some retention loss between successive unreinforced test trials that cannot be accounted for in terms of guessing shrinkage. The theoretical interpretation of this retention loss will be the subject of the next experiment.

Experiment II

Two principal questions arise concerning the retention losses observed in Exp. I. First, do the retention losses occur on an all-or-none basis, or do they represent reductions in correct response probability to intermediate degrees? Second, in either case, do the observed changes in response probability signify actual forgetting, either partial or complete, of stimulus-response associations established by preceding reinforcement? Information relevant to the first question can be obtained from a simple extension of the design used in Exp. I. It is necessary only to give at least three successive unreinforced test trials following the last reinforcement. The critical datum will be the proportion of instances in which items correct on the first such test but incorrect on the second are again correct on the third. If retention losses are all-or-none in character, this proportion should be no greater than chance. The second question is not so easily answered. An attempt will be made to secure some pertinent evidence by arranging conditions so that there is minimal opportunity for the learning of interfering responses between test trials and determining whether, under these circumstances, there is any progressive loss of information over the test series.

Method

The apparatus, source of *Ss*, and general procedure were the same as in Exp. I. The design shown in Table 3 was replicated with 12 different 8-item lists for each of 20 *Ss*, two items in each list being assigned to each

TABLE 3

DESIGN OF EXP. II IN TERMS OF REINFORCEMENTS (R) AND TESTS (T) FOR EACH ITEM TYPE ON EACH TRIAL

Item Type	Trial 1	Trial 2	Trial 3	Trial 4
1*i*	R T	T	T	T
1*d*	R	T	T	T
2*i*	R T	R T	T	T
2*d*	R	R T	T	T

row of the table. Item types labeled 1 and 2 received one and two reinforcements, respectively. The types labeled *i* (immediate) were tested immediately following the first reinforced cycle; those labeled *d* (delayed) were first tested on Trial 2. Thus items of Type 1*i* were reinforced only on Trial 1 but tested on all four trials, and so on.

Besides permitting evaluations of the proportions of items correct on a later trial after being first correct and subsequently not correct (henceforth denoted $C_n{:}N_{n-1}C_{n-2}$), the design provides for controlled comparisons relative to retention as a function of number of reinforcements and to retention as a function of tests interspersed between reinforcements. Thus it should be possible to differentiate the effects of two variables that have been confounded in standard experimental designs.

The stimuli were again sets of three consonant syllables, randomly drawn except for the elimination of duplicate syllables. For six lists the responses were the numbers 1 to 8; for the remaining six lists the responses were familiar one-syllable words drawn from an articulation test. The two types of responses were used in order to provide a basis for assessing the role of guessing on unlearned items.

Each *S* was trained and tested on six lists in each of two experimental sessions, number and word lists alternating within each session. Trial 1 on a list began with a reinforced cycle on which all eight stimulus-response pairs were exposed, then there was a test cycle with the stimuli belonging to items of Types 1*i* and 2*i* exposed alone. Trial 2 began with exposure of the stimulus-response pairs for items of Types 2*i* and 2*d*; then all eight stimulus members were exposed alone. On Trials 3 and 4, all eight items were retested, in a new random order each time, without intervening reinforcement.

Results

Response shifts.—Perhaps the simplest measure of retention is the probability with which a correct response to a given item on one test trial is followed by a correct response to that item on a later test (by the same *S*, of course) when no reinforcements have intervened. From the upper portion of Table 4, it is apparent that retention, so measured, was about 90% for adjacent pairs of unreinforced test trials, with no appreciable differences between items with words and items with numbers as responses. There appear to be no differences among item types in amount of retention over the last three test trials. For all types there is a consistent, though small,[3] increase in size of the conditional proportions over successive pairs of trials. Two possible interpretations of this trend im-

TABLE 4

PROPORTIONS OF REPEATED CORRECT RESPONSES ($C_{n+1}{:}C_n$) AND REPEATED INCORRECT RESPONSES ($I_{n+1}{:}I_n$) FOR EACH ITEM TYPE IN EXP. II

Item Type	Word Responses			Number Responses		
	$C_2{:}C_1$	$C_3{:}C_2$	$C_4{:}C_3$	$C_2{:}C_1$	$C_3{:}C_2$	$C_4{:}C_3$
1*i*	.74	.93	.93	.83	.90	.91
1*d*		.94	.92		.88	.93
2*i*		.92	.95		.89	.91
2*d*		.89	.93		.88	.95
Item Type	$I_2{:}I_1$	$I_3{:}I_2$	$I_4{:}I_3$	$I_2{:}I_1$	$I_3{:}I_2$	$I_4{:}I_3$
1*i*	.19	.27	.41	.17	.43	.47
1*d*		.30	.58		.41	.41
2*i*		.33	.24		.40	.39
2*d*		.13	.38		.16	.20

[3] The larger increases from $C_2{:}C_1$ to $C_3{:}C_2$ and from $I_2{:}I_1$ to $I_3{:}I_2$ may be attributable simply to changes in guessing probabilities consequent on the learning of additional Type 2 items on the second reinforced trial.

mediately suggest themselves. The first is that some of the stimulus-response associations established by the reinforced trials are, for some reason, more susceptible to forgetting than others; such differences would permit a selective factor to operate, the items whose responses are more resistant to forgetting being more heavily represented in the proportions for later pairs of trials. The second interpretation is that some learning may have occurred on test trials; other evidence bearing on this possibility will be presented in the next section.

Proportions of instances in which an incorrect response (excluding omissions) to a particular item was repeated from one trial to the next are presented in the lower portion of Table 4. Increases in these proportions from the second to the third pair of test trials[3] are larger, particularly for the items with word responses, than in the case of repeated correct responses. Again, the observed trend is compatible with the hypothesis that some learning occurs on test trials ("learning" being defined in terms of an increase in probability of whatever response actually occurs to a given stimulus), but possible artifacts cannot be entirely ruled out.

Proportions of shifts from incorrect responses on one trial to correct responses on the next (when there has been no intervening reinforcement) are shown in Table 5. The notation I_n has been used for incorrect responses *excluding* omissions on Trial n and is to be distinguished from N_n, used elsewhere for incorrect responses including omissions, and O_n, used in Table 5 to denote omissions. Thus $C_2:I_1$ represents instances in which *S*s made incorrect responses to particular items on Trial 1 and then made correct responses to the same items on Trial 2. As in Exp. I, the proportions of correct following incorrect are uniformly very much smaller than those of correct following correct responses. The fact that the proportions for both $C_2:C_1$ and $C_2:I_1$ are somewhat higher than in Exp. I is not unexpected. In Exp. II, each *S* learned 12 different lists; therefore, most of the data come from lists given after the initial period of "learning to learn" and after *S* has had an opportunity to develop relatively efficient habits of guessing on unlearned items. Allowing for this procedural difference, the results of the two experiments appear quite compatible. All values of $C_n:I_{n-1}$ in Table 4 fall well within the limit of what could be achieved by reasonably efficient guessing on unlearned items. The negligible differences between $C_{n+1}:I_n$ and $C_{n+1}:O_n$, together with the fact that each of these proportions is much larger for numbers than for word responses, would seem to support the interpretation that correct responses to previously noncorrect items are attributable to guessing. It would clearly be desirable to eliminate, or at least drastically reduce, the role of guessing on unlearned items; this is being attempted in experiments now in progress.

TABLE 5

PROPORTIONS OF CORRECT RESPONSES FOLLOWING CORRECT RESPONSES ($C_{n+1}:C_n$), INCORRECT RESPONSES ($C_{n+1}:I_n$) OR OMISSIONS ($C_{n+1}:O_n$) AND PROPORTIONS OF REPEATED INCORRECT RESPONSES ($I_{n+1}:I_n$) POOLED OVER ITEM TYPES FOR TEST TRIALS 2–4 OF EXP. II

	Word Responses	Number Responses
$C_{n+1}:C_n$	.92	.91
$C_{n+1}:I_n$	.13	.23
$C_{n+1}:O_n$	.11	.22
$I_{n+1}:I_n$	.34	.37

In order to determine whether retention losses occur on an all-or-none basis for individual *S*s we may examine the proportions for $C_n : N_{n-1}C_{n-2}$. If the occurrence of a noncorrect response to an item that *S* has responded to correctly on a previous test signifies a total loss of the correct association, then the probability of correct response to the given item on later tests, assuming no intervening reinforcement, should be no greater than that attainable by a random selection from among the unlearned responses. A substantial number of cases can be obtained by including both incorrect responses and omissions on Trial $n - 1$ and pooling over all four item types (using the first three test trials in each case). The proportion of .44, based on 126 observations, thus computed is enough larger than the simple proportions of correct following noncorrect shown in Table 5 to contraindicate the hypothesis of "total forgetting." One might raise the question whether the higher proportions of correct following noncorrect for items that have previously been responded to correctly is interpretable simply as a sequential effect; other things equal, one would expect that a response that has been given to an item on an earlier test would have a greater than chance probability on later tests for the given *S* even if the response had never been reinforced. As a control for such effects, a computation was made of the proportion of cases in which a particular incorrect response occurred to an item on a given trial after having occurred on the second preceding trial but not on the immediately preceding trial, i.e., $I_n : J_{n-1}I_{n-2}$, where I is a particular incorrect response and J is a different response to the same item. This proportion, based on 270 observations, proved to be .10. Apparently we must conclude that when a correct response to an item on one trial has been followed by an incorrect response on a later trial in a series of unreinforced tests, the probability of the correct response for the given *S* has been reduced to a value intermediate between that obtaining immediately after reinforcement and a chance level. In other words, retention losses do not necessarily occur on an all-or-none basis.

Proportion correct per test trial.—At a phenotypic level, partial retention losses clearly occur over a series of unreinforced tests in this situation. But the question remains whether these changes in observed response probability represent forgetting, i.e., weakening of the stimulus-response associations established by reinforcement, or the operation of some extraneous factor which affects performance but not learning. Evidence pertinent to this point is available in the over-all (unconditional) proportions of correct responses per test trial. If a progressive reduction in these proportions is observed, the interpretation will remain uncertain, for a reduction could be produced either by weakening of associations or by extraneous factors. If no reduction occurs, however, then certainly there has been no weakening of associations.

In Table 6 the proportions correct per test trial are given separately for the four item types and for word and number responses, each value being based on 240 observations. If one looks across the row of proportions for the four tests in the case of Type 1*i* items and the last three tests in the case of the other types, the salient feature is the complete absence of any tendency toward progressive reduction in proportions of correct

TABLE 6

PROPORTION CORRECT PER TEST TRIAL FOR EACH ITEM AND RESPONSE TYPE IN EXP. II

Item Type	Word Responses				Number Responses			
	Test Trials				Test Trials			
	1	2	3	4	1	2	3	4
1*i*	.41	.36	.40	.39	.55	.56	.60	.61
1*d*		.28	.32	.31		.49	.53	.58
2*i*	.43	.77	.78	.75	.58	.83	.78	.77
2*d*		.70	.67	.66		.73	.73	.78

responses from test to test. The only systematic trend apparent in the table is the progressive *increase* in proportions of correct responses from test to test in the case of the items with number responses and a single reinforcement. This effect, if significant, is probably attributable to increases in guessing efficiency. The frequency of omissions decreases by nearly 50% from Test 2 to Test 4 for these items, and consequently the contribution from guessing on unlearned items must be expected to increase also. To eliminate this factor, correct response proportions may be recalculated, based only on items which had either correct or incorrect responses, excluding omissions. When this is done the pooled proportions on Tests 2 through 4 run .62, .65, .64 for once-reinforced items and .84, .84, .83 for twice-reinforced items. Clearly there is no loss of information from the response protocols over the series of tests. Since no information is being fed into the system either, the conclusion seems inescapable that there is no progressive weakening of learned associations. The most parsimonious hypothesis as to the source of the response shifts described in the preceding section would seem to be one attributing them to fluctuations in the stimulus context from series to series.

Retention as a function of preceding reinforcements and tests.—Our design provides one fully controlled comparison relative to retention as a function of number of reinforcements, namely that involving item Type 1*d* vs. Type 2*d* on Tests 2 through 4. Examination of the proportions $C_3:C_2$ and $C_4:C_3$ in Table 4 reveals that, with ordinal number of the test trial and number of preceding tests equated, the difference between one and two preceding reinforcements has produced no mean difference whatever between these proportions for the two item types. Contrasting this result with the large difference between corresponding rows in Table 6, we may conclude that the effect of the second reinforcement upon retention was, if not actually zero, at least of a much smaller order of magnitude than its effect upon the amount learned.

Two ways are available to measure the effect of a test trial upon subsequent retention. The first is to compare proportions of correct responses on Test 2 for item Type 1*i* vs. Type 1*d* and for Type 2*i* vs. Type 2*d*. In each case, as may be seen in Table 6, there is an advantage of 5 to 10 percentage points for items that had a test immediately after the first, and preceding the second, reinforcement. Application of a rank test for paired replicates shows the differences to be significant at better than the .02 level for both comparisons. The second method, comparing proportions of $C_3:C_2$ and $C_4:C_3$ for Type 1*i* vs. 1*d* and 2*i* vs. 2*d*, shows only negligible differences (Table 4).

DISCUSSION

The bearing of our results upon extant theories requires no extended discussion.

The gist of the matter is that no hitherto published theory with which we are familiar gives a reasonable account of our principal findings. Assuming that possible artifacts have been adequately handled by our various control measures, the results on acquisition appear incompatible with any theory which expresses learning in terms of increments in associative strength, excitatory potential, or simply response probability in individual *S*s. The "pattern model" (Estes, 1959b) accounts for initial acquisition in this situation, but it is incomplete in that it does not provide for the observed retention losses from training to test trials or the pattern of response shifts occurring over a series of unreinforced test trials.

The shortcomings of the pattern model may be due in part to an unnecessary simplifying assumption. For convenience in calculations, when applying the model to the data of Exp. I, we assumed that with each consonant syllable there was associated a single stimulus pattern which recurred intact from trial to trial. However, there must be some change in the pattern between training and test trials, if only because two windows in the stimulus panel are illuminated in the former instance and only one in the latter. A minimal extension of the pattern model as originally published would include the assumption, common to all stimulus-response learning theories, that probability of a response reinforced on a training trial is reduced on subsequent test trials if components of the training stimulus pattern, including contextual, or "background," cues, have been replaced by novel ones in the test pattern.

An additional factor demonstrated in Exp. II is the occurrence of learning on the unreinforced test trials. It appears that the constancy of mean proportions of correct responses over the test series, the pattern of response shifts, and the increase in stereotypy over the test series would all be predictable on the assumption that the response occurring, whether correct or incorrect, on a test trial becomes conditioned by contiguity to the full pattern of cues present on the trial.[4] Formal presentation of the augmented pattern model for paired-associate learning suggested by the present findings will be deferred until independent empirical tests have been obtained from studies now in progress.

Summary

In Exp. I, 48 *S*s were given a single reinforcement (paired presentation of stimulus and response members) followed by two unreinforced tests on a set of eight paired-associate items. The relative frequency of correct responses on Test 2 for items having correct responses on Test 1 was approximately .7, whereas the relative frequency of correct following noncorrect responses was less than the .125 attainable by chance.

In Exp. II, the following design was replicated with 12 different 8-item lists of paired-associates for each of 20 *S*s: equal numbers of items received one and two reinforcements, with and without an interspersed test trial; then all items appeared on three unreinforced test trials. The mean proportion of correct responses remained constant over the series of tests following the last reinforcement; at the same time, significant proportions of shifts from correct to incorrect and vice versa occurred from test to test, but with the degree of stereotypy tending to increase progressively over the series. When other factors were controlled, retention was found to increase with the number of preceding tests but not with the number of preceding reinforcements.

The results were taken to support the view that acquisition occurs on an all-or-none basis, repeated reinforcements simply giving repeated opportunities for the formation of an association between a stimulus pattern and the reinforced response. A provisional interpretation of the data on retention involved two additional assumptions: that decrements in probability of the reinforced response are produced by changes in the stimulus context from training to test trials, and that the response, correct or incorrect, occurring on a test trial has some probability of becoming

[4] Judging by Guthrie's (1959) last publication on learning theory, this assumption comes closer to his intended interpretation of the contiguity principle than earlier formulations of statistical learning theory stated in terms of the independent conditioning of component cues.

conditioned to the stimulus pattern present on the trial.

References

Bourne, L. E., Jr., & Restle, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–96.

Bush, R. R., & Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 380–491. (a)

Estes, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Pp. 9–52. (b)

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

Guthrie, E. R. Association by contiguity. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 158–195.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.

Luce, R. D. *Individual choice behavior.* New York: Wiley, 1959.

McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* (Rev. ed.) New York: Longmans, Green, 1952.

Rock, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–93.

Spence, K. H. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Underwood, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, **51**, 276–82.

(Received November 6, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 340–348

EFFECTS OF TASK DEFINITION AND PROBABILITY OF REINFORCEMENT UPON THE ACQUISITION AND EXTINCTION OF IMITATIVE RESPONSES[1]

VERA T. KANAREFF AND JOHN T. LANZETTA

Fels Group Dynamics Center, University of Delaware

In a previous experiment (Kanareff & Lanzetta, 1958) the effectiveness of the probability of reinforcement in modifying the level of imitation was found to vary with the prevailing social sanctions regarding its use. Apparently anticipation of disapproval for imitation inhibited the use of imitative responses even though they were instrumental to goal attainment. The present study tests the generality of these findings by imposing the imitation paradigm on a different task and by ascribing the social sanctions to another source. It also provides an opportunity for examining the imitative response under conditions of "extinction." With data on the extinction of imitation and other social responses so meager at the present time, it is not known whether or not the parameters governing the extinction of socially neutral responses would operate similarly on a social response such as imitation.

The negative and positive sanctions in the previous studies were induced through instructions stressing either the desirability of separate judgments or the facilitating effect of working in a group, respectively. The perceived source of sanctions may well have been the experimenter and *Ss* may have performed as they thought the experimenter wished them to perform. It was thus desirable to replicate the findings in situations where expectations of approval were likely to be already well established through the socialization process and independent of the experimenter.

Although imitation may have been strongly and indiscriminately reinforced early in his development, an individual must eventually learn to discriminate between those situations in which imitation is legitimate and those in which it is not. Since feedback as to correctness of the discriminations is likely to be probabilistic and since the discriminations should generalize to other situations differing in some degree on one or more dimensions from the situations he has encountered, the individual may be required rather to form judgments of the likelihood that imitation would be legitimate in a given social situation and classify social situations on the basis of such judgments.

One situation which was expected to elicit low expectations of the legitimacy of imitation for college students was a test of problem-solving ability. Schein (1954) attributed the low level of imitation among his adults, in part, to their compunctions about cheating (imitating) on a test and being caught at it. Even when a stooge gave objectively correct

[1] This research was supported in part by the United States Air Force under Contract AF 33(616)-5845, monitored by Aero Medical Laboratory, Directorate of Research, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio.

The writers wish to express their appreciation to Jo Ann Davis for her assistance in collecting and analyzing the data, and to Juanita Randolph for her assistance in analyzing the data.

answers on a visual test (allegedly part of a new IQ test being tried out by the Board of Education) Luchins (1944) found that not all children would imitate the stooge's responses. Apparently, concern about cheating (imitating) and getting into trouble for it was aroused. Thus it seemed that a comparison between test-like situations and less achievement-oriented or ego-oriented situations would be desirable. The situation selected for comparison was a gambling or guessing one, for imitation was expected to be more socially acceptable when chance rather than ability determined the outcomes.

A further motivation for employing these particular situations was that comparable data on a socially neutral response is available. Goodnow (1955) found a higher concentration of choices of the more frequently occurring event for a gambling task as compared with a problem-solving task in a Humphreys'-type guessing situation. Both Goodnow's findings and the above rationale lead to the same prediction: when imitation is instrumental, imitative responses will be more frequent under the gambling definition of the task than under the problem-solving definition. However, no differences between the two task definitions are expected when imitation is not instrumental to success.

Method

General Procedure

The procedure was a modification of previously used procedures (e.g., Lanzetta & Kanareff, 1959). Three *Ss*, seated in separate, partially enclosed booths preventing intercommunication, predicted whether a red or green light would flash on. Each *S* indicated his prediction by pressing the appropriate key on the response box, a small metal box containing two response keys and two pairs of signal lights. The three *Ss* were told they would give their predictions in a fixed order and that the judgment of one of the two other persons would be indicated by the pair of lights directly above his response keys. Since all three *Ss* were instructed that their partner would give his judgment first and since the signal lights in the booths were all connected on a central control console, *E* could present a programmed sequence of "partner's" judgments to each *S*.

The correctness of *S*'s choice was indicated by the top pair of lights (red and green) on the response box, the sequence of these lights also being prearranged.

Experimental Conditions

Probability of reinforcement.—The sequence of events to be predicted and the corresponding two response sequences for the partner were randomized within blocks of 10 trials with the restriction that there be an equal number of "red" and "green" predictions. Where equal distribution of the two choices was not possible the odd choices were balanced over blocks of trials. The two partner's response sequences formed the two probabilities of reinforcement, .8 and .5.

The initial 100 trials constituted the acquisition period and the remaining 50 trials, the extinction period. During acquisition, 80% of the partner's predictions were correct in the .8 reinforcement condition and 50% of them were correct in the .5 reinforcement condition. During extinction, 50% of the partner's predictions were correct for both reinforcement conditions.

Definition of task.—For half the *Ss* the task was defined as a problem-solving test. The relevant portion of the instructions is as follows:

> There has been a growing interest in the field of psychology in tests of problem-solving behavior. We, in particular, are interested in testing your ability to solve problems when clues are not readily available. In this test we will see how accurately you can predict which of two lights will flash on. . . . Although this is a difficult problem-solving test, most of our subjects get at least 80% of their predictions correct. . . . Remember, this is a test of problem-solving behavior, and we are interested in testing your ability to solve problems where clues are not readily available. In this test we will see how accurately you can predict which of two lights will flash on. Although this is a difficult

problem-solving test most of our *Ss* do get at least 80% of the predictions correct.

The task was defined as a gambling game to the rest of the *Ss*:

> There has been a growing interest in the field of psychology in guessing behavior. We, in particular, are interested in seeing how well people can do at guessing a series of things when they have no observable clues to guide them. In this study we will see how accurately you can predict which of two lights will flash on. . . . Although this is purely a guessing game, most of our *Ss* get at least 80% of their predictions correct. . . . Remember, this is a study of "guessing behavior" and we are interested in seeing how well people can do at guessing a series of things when they have no observable clues to guide them. We want to see how accurately you can predict which of two lights will flash on. Although this is a guessing game most of our *Ss* do get at least 80% of the predictions correct.

Experimental Design

The two probabilities of reinforcement and the two task descriptions form a 2 × 2 factorial design. Twenty-four male and 24 female University of Delaware summer student volunteers were randomly assigned in three-person groups of the same sex to the four cells of this design with the restriction that each cell contain an equal number of groups of each sex.

The number of imitative responses (judgments indicating agreement with the partner) per block of 10 trials is the major dependent variable. A questionnaire assessing *S*'s aspiration level and estimates of his success, his partner's success, and of the amount of agreement with the partner was also used to provide data about *S*'s expectations in the experimental situation.

Results

Imitative Responses

The prediction had been made that a higher level of imitation would obtain throughout the acquisition period for gambling than for problem solving when imitation was instrumental; no difference in level was expected

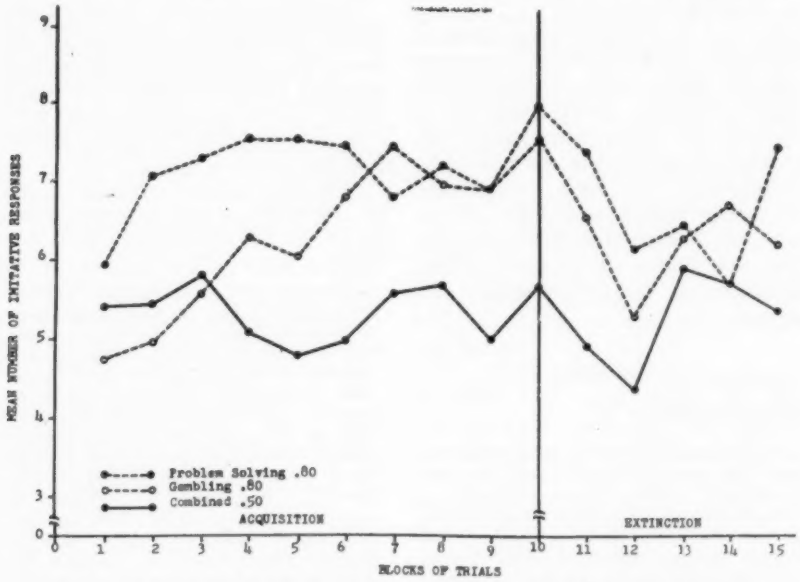

FIG. 1. Mean number of imitative responses per block of 10 trials as a function of task definition and probability of reinforcement.

TABLE 1

TREND ANALYSES OF IMITATIVE RESPONSES UNDER THE .5 AND .8 PROBABILITIES OF REINFORCEMENT

Source	*df*	Error Term	.5 Probability of Reinforcement		.8 Probability of Reinforcement	
			MS	*F*	*MS*	*F*
A. Overall trend	(9)	E	(2.96)	1.43	(11.46)	4.41**
1. Linear	1	E_1	.0409		78.40	14.18**
2. Quadratic	1	E_2	3.89	2.58	9.89	3.04
3. Cubic	1	E_3	.0688		8.97	4.65*
4. Deviation from	6	E_4	4.53	2.42*	.98	
B. Bet. grp. means	(3)	D	(10.36)		(40.15)	3.37*
a. Description	1	D	2.40		43.35	3.64
b. Sex	1	D	26.67	1.79	43.35	3.64
c. D × S	1	D	2.01		33.75	2.84
C. Bet. grp. trends	(27)	E	(2.06)	1.00	(3.12)	1.20
1. Linear	(3)	E_1	(.61)		(8.61)	1.56
a. D	1	E_1	.10		23.56	4.26
b. S	1	E_1	1.70		1.37	
c. D × S	1	E_1	.02		.89	
2. Quadratic	(3)	E_2	(.45)		(3.80)	1.17
a. D	1	E_2	.01		.10	
b. S	1	E_2	1.33		10.34	3.18
c. D × S	1	E_2	.01		.96	
3. Cubic	(3)	E_3	(2.06)	1.15	(3.45)	1.79
a. D	1	E_3	5.97	3.33	10.20	5.28*
b. S	1	E_3	.01		.13	
c. D × S	1	E_3	.19		.03	
4. Deviations	18	E_4	2.56	1.37	2.04	
D. Bet. ind. means	(20)	E	(14.92)	7.24**	(11.90)	4.58**
E. Bet. ind. trends	(180)	E	(2.06)		(2.60)	
1. Linear	20	E_1	4.02		5.53	
2. Quadratic	20	E_2	1.51		3.25	
3. Cubic	20	E_3	1.79		1.93	
4. Deviation from	120	E_4	1.87		2.12	
Total	239					

* $P = .05$.
** $P = .01$.

otherwise. Contrary to prediction, the problem-solving and gambling acquisition curves reached and fluctuated about a common level under the .8 reinforcement (see Fig. 1). The common level, however, was attained much more rapidly by *Ss* involved in problem solution. An orthogonal trend analysis (Grant, 1956) of the acquisition data in the .8 reinforcement condition indicates that the two curves differ significantly in their cubic components and almost significantly in their linear components. The summary of the trend analysis is given in Table 1.

When imitation was not instrumental to task achievement the prediction was upheld. In the orthogonal trend analysis of the .5 reinforcement condition, also summarized in Table 1, none of the description effects of the components, nor any of the components of the Between Groups Trend, nor the Between Groups Trend itself is significant.

A comparison of the two trend analyses indicates that the two acqui-

sition curves differ significantly in their linear and cubic components, the .8 reinforcement curve having significant linear, and cubic overall trends, and the .5 reinforcement curve having no significant overall trends (see Fig. 2). In both analyses large individual differences in response to the experimental manipulations are evident.

An analysis of variance over the acquisition trials of both probability of reinforcement conditions contributed little additional information but confirmed the earlier finding of a higher level of imitation under .8 reinforcement, the means per block of trials being 6.78 and 5.38 for the .8 and .5 reinforcements, respectively. The only additional finding is the significant Probability of Reinforcement × Sex interation. The Duncan range test (1955) results on the means are presented in Table 2. Both men and women imitated more in the .8 than in the .5 reinforcement condition, the difference being significant only for the men. Differences between the sexes were not significant in either of the probabilities of reinforcement. Men may be more concerned about being successful than about the legitimacy of the means employed, whereas

TABLE 2

IMITATION MEANS FOR PROBABILITY OF REINFORCEMENT × SEX*

Sex	Probability of Reinforcement	
	.80	.50
Men	7.20[a]	5.04[c]
Women	6.35[ab]	5.71[bc]

* Those means marked with a letter in common do not differ significantly at the .05 level according to the Duncan range test.

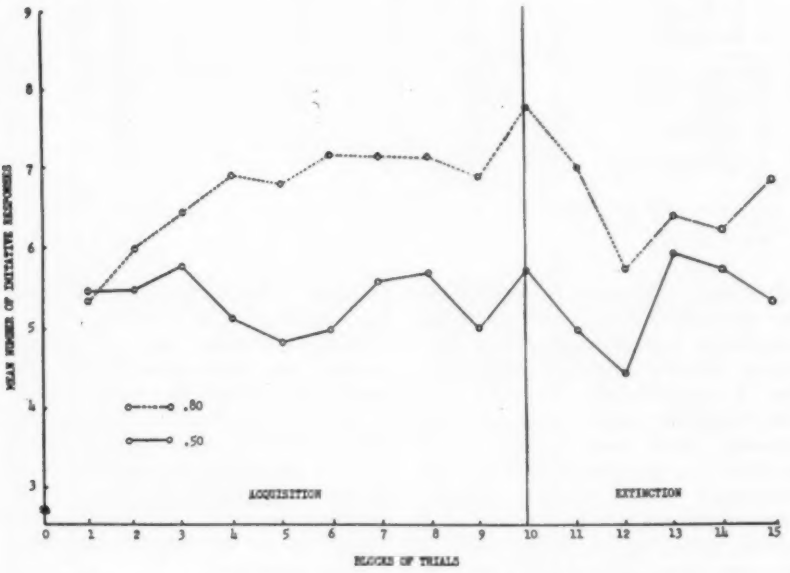

FIG. 2. Mean number of imitative responses during acquisition and extinction trial blocks for .8 and .5 probabilities of reinforcement.

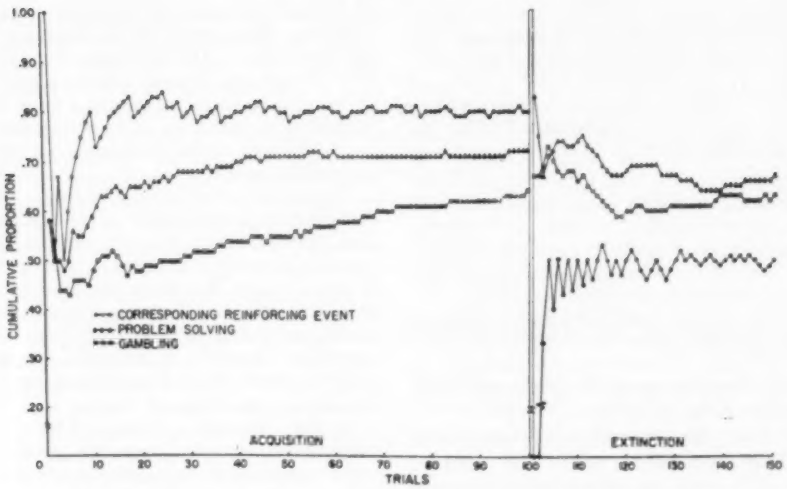

FIG. 3. Cumulative proportions of the imitative response and its reinforcing event under the .8 probability of reinforcement.

women may be more concerned about using legitimate means.

As in previous studies of imitation (Kanareff & Lanzetta, 1958; Lanzetta & Kanareff, 1959), the cumulative proportion of the instrumental imitative response in both .8 reinforcement conditions approaches but remains below the cumulative proportion of the corresponding reinforcing event throughout acquisition (see Fig. 3). The extent of undermatching again is determined by the experimental conditions. Although the level of imitation under the .8 reinforcement does drop during the initial part of the extinction period,[2] it rises again later (Fig. 2). No significant differences were found between gambling and problem solving during extinction.

[2] The term extinction is not used in the usual sense of complete withdrawal of all reinforcement nor as a shift to continuous reinforcement of the other response alternative.

Questionnaire Responses

In general, *S*s overestimated their partner's successes when the partner was correct 50% of the time, estimating on the average 59.17% correct, and gave significantly more accurate estimates when he was correct 80% and then 50% of the time, estimating on the average 71.25% correct. Similar results were obtained in previous studies (Kanareff & Lanzetta, 1958; Lanzetta & Kanareff, 1959). However, as indicated by the significance of the Probability of Reinforcement × Task Definition × Sex interaction in the analysis of variance of these deviation scores, the relationship obtained only for the men engaged in the gambling task. The Duncan range test results on the appropriate means are given in Table 3. With the exception of the men with .8 probability of reinforcement, the gambling *S*s overestimate the partner's performance

TABLE 3

DEVIATION MEANS FOR PROBABILITY OF REINFORCEMENT × TASK DEFINITION × SEX*

Task Definition	Sex	Probability of Reinforcement	
		.80	.50
Problem Solving	M	−13.34[c]	8.33[a]
	F	3.33[b]	−3.33[c]
Gambling	M	−1.67[c]	−8.33[d]
	F	−12.50[de]	−11.17[de]

* Those means marked with a letter in common do not differ significantly at the .05 level according to the Duncan range test.

significantly more than the problem-solving *Ss*.

The *Ss* generally underestimate their agreement with the partner significantly more when the task is described as a gambling task (a deviation of 10.83%) than when it is described as a problem solving one (a deviation of 2.00%); also men underestimate the amount of agreement more than women (12.21% as compared with .62%).

There were no significant findings other than individual differences in the remaining analyses of the questionnaire data.

DISCUSSION

Several earlier findings concerning the effects of probability of reinforcement of imitation on imitation in the judgment of tones were confirmed in a verbal conditioning situation: (*a*) A relatively typical acquisition curve results from the application of a .8 reinforcement schedule whereas a .5 schedule yields a curve which wanders about the event matching value. (*b*) Undermatching of the reinforcing event occurs throughout acquisition when imitation is instrumental, the extent of undermatching being determined by the experimental conditions.

Goodnow's findings (1955) and our theoretical rationale had led us to expect more rapid acquisition and a higher level of imitation under the gambling task definition than under the problem-solving definition. In fact, however, the level of imitation in early trials was much higher for problem solving, the level of imitation approaching a level common to both task definitions later in the acquisition period.

Apparently, Goodnow's findings are not directly generalizable to the imitation paradigm, nor does the legitimacy variable appear to function as predicted. A reanalysis of the imitation paradigm as a discrimination learning experiment would point out a basic difference in procedure between Goodnow's study and the present one and at the same time, account for the obtained results.

In the imitation paradigm the discriminative stimuli (partner's predictions) are reliable (correctly anticipate the event to be predicted) on a given proportion of the trials and is reversed on the remaining proportion. In the .5 reinforcement condition, both discriminative stimuli are followed by each event with a .5 probability; in the .8 reinforcement condition S_1 is followed by E_1 and S_2 by E_2 with a .8 probability while the reverse relationship between partner's predictions and events to be predicted occurs with a .2 probability.

Two predictions follow immediately from statistical learning theory and experimental findings on discrimination learning (e.g., Estes & Burke, 1955). The mean learning curve for the .5 reinforcement condition should vary around five imitative responses on each block of trials during acquisition and extinction. The mean learning curve for the .8 reinforcement condition should approach an asymptote of eight imitative responses. With only 100 acquisition trials, however, and with discrimination learning proceeding more slowly than simple verbal learning, and "social" discrimination learning proceeding more slowly than nonsocial discrimination learning (Neimark, 1959), undermatching of the asymptotic value should hold throughout acquisition. During extinction the mean .8 reinforcement learning curve should drop slowly toward five

imitative responses, as the new discrimination is being learned, or rather, the old one, unlearned. The data are consistent with these predictions.

The initial tendency of making an observing response, any response resulting in exposure to the pair of discriminative stimuli involved in discrimination learning (Wyckoff, 1952), should be higher when the task is defined as problem solving and *S*s are instructed that clues are not readily available than when it is defined as gambling and *S*s are instructed that no observable clues are present. Consequently when the discriminative stimuli have better than chance reliability, the gambling group should learn the discrimination more slowly than the problem-solving group and the difference between them should be especially marked on the early trials. When the discriminative stimuli have only chance reliability no differences between the two groups is anticipated. Again the data are in accord with the predictions.

Although the discrimination learning analysis provides an excellent account of the imitative response, it does not account for the tendency of the gambling group to overestimate the reliability of the discriminative stimuli and underestimate its utilization of the discriminative stimuli significantly more than the problem-solving group, nor does it account for the experimental results of the earlier study (Kanareff & Lanzetta, 1958). Consideration of the prevailing sanctions towards the *utilization* of discriminative cues would seem especially vital for social discrimination learning where the discriminations based on information about another persons' behavior are more likely not to be utilized in full, or at all, in spite of a high occurrence of observing responses and a high degree of discrimination (e.g., Luchins, 1944; Neimark, 1959; Schein, 1954).

In the present study the experimental manipulations of the expectations of social approval and disapproval for imitation may have simultaneously affected the utility of task success and failure or the utility of social approval and disapproval as well as the probability of occurrence of the observing response. The contamination of the experimental manipulations might then account for the failure of the legitimacy variable to operate as predicted. Another possibility is that the utilization of information about another person's behavior to learn a principle or as a result of applying a principle in problem solution may be considered to be legitimate behavior. Only when such information is used simply to get a high score would it be perceived as less appropriate in problem solution than in gambling.[3] The importance of the prevailing sanctions towards the utilization of discriminative cues in discrimination learning is an issue that obviously requires further empirical determination.

Summary

The level of imitation was previously found to be a function of both its probability of reinforcement and the prevailing social sanctions (Kanareff & Lanzetta, 1958). The present study tests the generality of these findings to another task and another source of sanctions, and also examines the imitative response under conditions of "extinction."

Several of the earlier findings were confirmed: (*a*) A relatively typical acquisition curve results under a .8 reinforcement schedule whereas a .5 reinforcement schedule fluctuates about five imitative responses on each block of trials. (*b*) The extent of undermatching of the reinforcing event when imitation is the instrumental response is determined by the experimental conditions.

The predicted difference in the level and rate of acquisition of imitation between the gambling and problem-solving task definitions, definitions which were expected to arouse differing expectations as to the legitimacy of imitation, did not obtain. The level of imitation was much higher for problem solving in the early trials; later in the acquisition period it reached a level common to both task definitions.

[3] The differentiation between cognitive and rote utilization of "social" discriminative stimuli is an extension of the distinction made by Luchins (1944) between cognitive and rote imitation.

A reanalysis of the imitation paradigm as a discrimination learning paradigm provides an excellent account of the present set of results but statistical learning theory and empirical findings on rate of acquisition of discrimination learning do not consider the possibility that information about another person's behavior may not be utilized in spite of a high occurrence of observing responses and a high degree of discrimination.

References

Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–41.

Estes, W. K., & Burke, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.*, 1955, **50**, 81–88.

Goodnow, J. J. Determinants of choice distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Kanareff, V. T., & Lanzetta, J. T. The acquisition of imitative and opposition responses under two conditions of instruction-induced set. *J. exp. Psychol.*, 1958, **56**, 516–528.

Lanzetta, J. T., & Kanareff, V. T. The effects of a monetary reward on the acquisition of an imitative response. *J. abnorm. soc. Psychol.*, 1959, **59**, 120–127.

Luchins, A. S. On agreement with another's judgment. *J. abnorm. soc. Psychol.*, 1944, **39**, 97–111.

Neimark, E. D., & Rosenberg, S. The effect of "social" discriminative cues on probability learning. *J. exp. Psychol.*, 1959, **58**, 302–311.

Schein, E. H. The effect of reward on adult imitative behavior. *J. abnorm. soc. Psychol.*, 1954, **49**, 389–395.

Wyckoff, L. B. The role of observing responses in discrimination learning. Part I. *Psychol. Rev.*, 1952, **59**, 431–442.

(Early publication received February 1, 1960)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 349–358

ACQUIRED DRIVE STRENGTH AS A JOINT FUNCTION OF SHOCK INTENSITY AND NUMBER OF ACQUISITION TRIALS[1]

MELVIN L. GOLDSTEIN

University of Illinois

Empirical laws describing the growth and decay of aversive responses have not been established with the degree of generality implied by such theoretical constructs as "acquired drive," "fear," and "anxiety." This state of affairs is attributable to the use of diverse conditioning procedures without sufficient consideration of their complex effects on criterion responses. The present study was therefore designed to explore the joint effect of several magnitudes of shock intensity and acquisition trials in a *relatively simple* aversive conditioning situation.

The acquired drive paradigm consists of two phases: a "classical" acquisition session during which a neutral stimulus is paired with a UCS of *fixed duration*, and a test session that measures the effects of CS presentation in the absence of shock. The test may require the acquisition of new responses that are motivated by CS onset and reinforced by CS termination (Brown & Jacobs, 1949; Kalish, 1954), or it may require performance of previously learned (Mathers, 1957) or "innate" (Brown, Kalish, & Farber, 1951) response sequences. Acquired drive strength, in such experiments, is assumed to be *directly* related to strength of the test response. Modifications of the test procedure that emphasize the development of competing responses (Dinsmoor & Campbell, 1956a, 1956b) or of inhibitory processes (Libby, 1951) yield test responses that are weakened by CS onset and presumably reflect an *inverse* relation between the dependent variable and acquired drive strength.

The present experiment used a factorial design and Kalish's (1954) acquired drive method to investigate the aversive properties of CS as a joint function of four CS-UCS pairing levels and three intensities of UCS; the levels selected for both variables covered a broad range of magnitudes. It was predicted that, under these conditions, the function relating strength of a hurdle-jump test response to the number of CS-UCS pairings would be different at each shock intensity level and that a significant interaction would be found between the two independent variables. It was suggested that the three functions might reach the same maxima at rates varying directly with shock intensity.

METHOD

Experimental Design

A $3 \times 4 \times 10$ mixed factorial design (Lindquist, 1953, pp. 281–284) was used in order to facilitate the analysis of repeated measurements of the test response and to test the hypothesis that the two independent variables interact in their effect on performance. The three columns and four rows of the design corresponded, respectively, to the levels of shock intensity and number of CS-UCS pairings: 100-, 150-, and 790-v. ac delivered through a fixed resistance of

[1] Based on a dissertation submitted to the Graduate College of the University of Illinois in partial fulfillment of the requirements for the PhD degree. G. Robert Grice directed the study and contributed to the solution of numerous problems that arose during the investigation.

250,000 ohms; and 1, 3, 9, and 20 "classical" CS-UCS presentations. The third dimension of the design represented the number of hurdle-jump test trials (ten 3-trial blocks) and was the same for all *Ss*. The design therefore contained 12 independent experimental groups.

Eight *Ss* were assigned at random to each of the 12 treatment-combinations immediately before the start of the experiment. Four *Ss* belonging to the same group were given their appropriate number of CS-UCS pairings, at the proper shock level, simultaneously during the classical conditioning session. The order of running was randomized for the squads of 4 *Ss* and was maintained unchanged throughout the experiment. The *Ss* within each squad were run in random order on the first trial of the hurdle-jump test session and in the same sequence on subsequent test trials.

A control group of 8 *Ss* was given 20 backward conditioning trials at the 790-v. shock level. The *Ss* were assigned at random to this group but the order of running was not randomized: 4 control *Ss* were run during the middle of the study, after data had been collected for 48 experimental *Ss*, and 4 were run at the end. Treatment of the control group during hurdle-jump test trials was identical with treatment of experimental *Ss*.

Subjects

The *Ss* were 104 experimentally naive male albino rats, supplied by the Holtzman Company, Madison, Wisconsin, and housed in the animal room maintained by the psychology department of the University of Illinois. They were 94 days old on arrival and 119–136 days old at the beginning of the experiment. Although *Ss* ranged in weight from 356 to 452 gm. at the start of the experiment, the mean weights of the 13 groups were restricted by the sampling procedure to 399.5–424.8 gm. There was no evidence of systematic differences among these means.

The *Ss* were adapted to the living conditions of the animal room, and to *E*, for a period of approximately 30 days after arrival. They were housed two *Ss* per cage in wire cages containing metal floors covered with wood-shavings and were permitted free access to tap water and Purina laboratory chow. They were tamed by handling for approximately 3 min. per day during the first 7 days of this pre-experimental period and were weighed every two or three days thereafter until the first experimental day. Overhead lights that illuminated the animal room were left on whenever *Ss* were handled or run.

The experiment proper was conducted between 8:30 P.M. and 8:30 A.M. Four *Ss* in the control group, however, were unavoidably run during the late morning and early afternoon.

Apparatus

The apparatus consisted of four identical boxes; each was divided by a guillotine-type door and a hurdle into two compartments of equal size that contained a grid floor in one partition and a wooden floor in the other. The compartments were 10 in. long, 5 in. wide, and 6½ in. high (inside dimensions). The hurdle was 2 in. high and was surmounted by a door that, when raised by *E*, presented a 4½-in. × 4½-in. opening through which *S* could jump from the grid to the wooden floor. The hurdle was made of thin gauge aluminum sheet. The grid floor was constructed of $\frac{3}{32}$-in. stainless steel rods spaced at ½-in. intervals. The back of the box was made of ¼-in. plywood, the two end panels were ¾-in. white pine, the door was ⅛-in. Masonite, and the front was ¼-in. Plexiglas. Separate hinged Plexiglas lids covered each compartment and were held securely in place by a removable Plexiglas clamp that prevented *S* from climbing out of the box. During the conditioning session an aluminum clamp held the door closed and prevented *S* from escaping into the nonelectrified compartment. All wooden parts of the box, as well as the hurdle, were painted with lead-free, flat black, enamel. A thin polyethylene film covering the front of the box served as a one-way screen. The boxes were arranged in a row on a table and were separated by approximately 2½ in. The area 3½ in. beneath the grid floor was covered with brown wrapping paper. Four black wooden carrying cages, of the same dimensions as the wood-floor compartments and covered with hinged hardware cloth lids, were used to transport *Ss* from the animal room to the experimental room.

The compound CS consisted of a series of buzzers and lights mounted on the grid-sides of the boxes. One buzzer (Edwards, No. 15, Size 2) was attached to the outside of each grid compartment at the center of the end panel and two light bulbs were supported, one directly over the other, by a black aluminum rod that was fastened to the same end panel. The upper light bulb

(115 v., 6 w., unfrosted) extended 20 in. above the grid floor; the lower bulb (115 v., 10 w., unfrosted) was 2 in. below the upper. Both bulbs protruded approximately 1¼ in. into the area over the grid compartment. The four upper lights provided a constant level of background illumination for the four boxes and for the experimental room. The levels of background illumination in the four grid compartments, measured with a Macbeth illuminometer as light reflected from the Standard test plate to a point approximately 3 in. above the center of the grid floor, ranged from .78 to .98 ft-c. The lower lights served as one component of the CS and, when turned on, increased the levels of illumination to 5.4–5.8 ft-c. An exhaust fan located in the ceiling of the experimental room was used as the source of a masking noise. Measurements made with a General Radio sound level meter Type 759-B showed that background noise levels in the four grid compartments ranged from 53.5 to 55 db on the "B" scale of the meter and from 68.5 to 69.5 db on the "C" scale. Activation of the four buzzers increased the noise levels to 83–84 db on the "B" scale and to 84–84.5 db on the "C" scale. All meter readings were made with the microphone of the instrument suspended approximately 3 in. above the center of a grid floor and with the lids of the other compartments closed. The buzzers and lights were always activated simultaneously and were interrupted at the rate of 2½ times per sec.

The UCS consisted of a 60-cycle ac current that was conducted to each grid through a 250,000-ohm series resistor. The four grids and four resistors were, in turn, connected in parallel, thereby enabling *E* to shock four *Ss* simultaneously. The network formed in this manner was connected to the secondary coil of a variable transformer capable of delivering voltages ranging from 0 to 1200 v. The primary coil was connected directly to the 115-v. line. A Simpson meter was incorporated into the circuit across the output of the transformer at a point preceding the grid-resistor network and was used to measure the three voltages employed in the experiment. Voltage measurements were made at the beginning and end of each conditioning session.

The durations of CS and UCS, and the interval between CS and UCS onset, were automatically controlled by a mechanical timing device. The intervals between successive CS-UCS presentations were timed with a stop watch. During the test session, the latency of each hurdle-jump response was measured to the nearest .01 sec. by a Standard Electric timer; the timer and CS were activated simultaneously by a microswitch that made contact with the guillotine-type door only when the latter was completely raised by *E* and broke contact when the door was lowered a fraction of an inch. This arrangement may have accounted for some of the extremely short latencies obtained in this study.

Procedure

Adaptation to apparatus (Day 1).—The running schedule was divided into four sessions conducted over a period of four successive days. The procedure on Day 1 was designed to adapt *Ss* to the apparatus. Four *Ss* belonging to the same group were removed from their living cages, one *S* at a time, and were placed directly into the black carrying cages to which they had previously been assigned at random. The carrying cages were put on a shelf in the animal room where they remained until each had been filled with one *S*. Two *Ss* were then taken to the experimental room and put on another shelf to wait until all four *Ss* had been transported. One *S* was then placed into each grid-compartment, facing the door, and remained there undisturbed for 1 hr. The *E* was not in the room during this hour, but *Ss* were prevented from leaving the box or entering the wood-floor compartment by the Plexiglas and aluminum clamps. The background lights were on throughout this session, but neither CS nor UCS was activated. At the end of the hour, *Ss* were removed from the grid compartments to the carrying cages and then were returned to the shelf in the experimental room. They were taken back to the shelf in the animal room two at a time and were put back into the living cages directly from the carrying cages.

Before each squad of *Ss* was run, the grids were wiped with rubbing alcohol and were permitted to dry for about 1½ hr. Grids were also wiped on the conditioning day, but not on the hurdle-jump test days.

Classical conditioning (Day 2).—During Day 2, the squad of four *Ss* was given the combination of CS-UCS pairings and UCS intensity appropriate to its group. This session was begun at approximately the same hour as the adaptation session of the preceding day. The *Ss* were transported to and from the experimental room and were placed into and taken out of the grid-compartment in the manner described above. The *E* was seated about 3 ft. in front and to the center of the row of boxes in order to facilitate

inspection of the grids for feces and to permit presentation of CS and UCS by remote control. The interval between CS and UCS onset was 4 sec., UCS duration was 1 sec., and CS and UCS were terminated simultaneously by the automatic timing system. Successive CS-UCS presentations were separated by 3-min. intervals. Since all *Ss* remained in the grid-compartment for 69 min., regardless of their group, the initial CS-UCS presentation occurred during different portions of the session for the 1-, 3-, 9-, and 20-trial groups. The *Ss* in the 1-trial group received their single CS-UCS pairing 66 min. after they had been placed into the grid-compartment; the remaining groups received their initial trials on the 60th, 42nd, and 9th min. of the session, respectively. All groups were removed from the grid-compartment 3 min. after the last trial and therefore had been exposed to background cues for equal periods of time. Preliminary observations of pilot *Ss* had indicated that many *Ss* went to sleep, under these conditions, before their first shock trial. Accordingly, *Ss* in the main experiment were roused 3 min. before their initial trial. The *E* accomplished this by opening the box and, whether *S* was asleep or not, momentarily lifting *S* approximately 2 in. from the grid floor.

Hurdle-jump test (Day 3).—The purpose of this session was to measure the effectiveness of the classical conditioning procedure. The *Ss* were required to learn to jump across a hurdle from the grid-floor into the wood-floor compartment when CS was activated and the door separating the two compartments was opened. The *Ss* were never shocked on this day and had never before been over the hurdle. Each *S* received 15 hurdle-jump acquisition trials during a 2-hr. period that began at about the same hour as the previous sessions. At the start of this session, the squad of four *Ss* was lined up in a row on the shelf in the animal room with each *S* in its proper carrying cage. The *E* then took the first *S* into the experimental room, placed the carrying cage in front of the box in which *S* had been conditioned, waited for 10 sec., then put *S* into the grid-compartment, facing toward the door, and secured the lids with the Plexiglas clamp. Ten seconds later, the door was quickly raised by *E*, thus simultaneously closing the microswitch and activating CS and the Standard Electric timer. The CS and the timer remained on until S had crossed over the hurdle into the opposite compartment. The *E* thereupon lowered the door, thus turning off both CS and the timer, and waited for 10 sec. before removing *S* from the wood-floor compartment to the carrying cage. The *S* was left in the carrying cage for an additional 10 sec. while *E* recorded the interval between CS onset and termination from the timer. This interval was designated as *S*'s "hurdle-jump latency" for the test trial and was the response measure used throughout the experiment. The *S* was then returned to the shelf in the animal room and remained there for 6 min. while *E* gave the other three members of the squad their first test trials in the above manner. This procedure was repeated until the squad had received 15 test trials with intervals of about 8 min. between CS presentations.

Maximum CS duration permitted on any hurdle-jump trial was 30 sec. If *S* had not crossed the hurdle by that time, *E* lowered the door and, 10 sec. later, removed *S* from the grid-compartment. Latencies of 30 sec. were recorded for such trials.

Second hurdle-jump test (Day 4).—The second test day was designed to insure development of maximum hurdle-jump response strength. This session was conducted 24 hr. after Day 3 with the identical procedure used during that session.

Control group.—The purpose of this group was to determine whether conditioning had occurred to stimuli other than the CS (buzzers and lights) specified by *E*. Adaptation and hurdle-jump training were the same as for experimental *Ss*, but the control group received 20 backward conditioning trials at the 790-v. level during the classical conditioning day. The interval between UCS termination and CS onset was 20 sec.; the durations of UCS and CS were 1 and 5 sec., respectively.

Results

The hurdle-jump latency recorded on each test trial was transformed to a reciprocal and then summed over blocks of three successive test trials for each *S* individually. The mean of a block of three reciprocals was the score used in a series of analyses of variance.

The original data analysis plan had provided for separate analyses of variance for Days 3 and 4. It had been anticipated that the functions relating the response measure to test trials would be maximized on the last block of Day 3 test trials, for all 12

experimental groups, and would show no decrement during Day 4. The predicted Shock Levels × Number of CS-UCS Pairings interaction was expected to be homogeneous throughout both days and identical in form at each test trial level. Tests of the significance of this interaction had been planned for scores obtained during the latter part of Day 3 and all of Day 4.

An analysis of variance utilizing data from the entire 10 blocks of test trials revealed, however, that there was a triple interaction, significant beyond the 5% level, and that a highly significant interaction existed between shock levels and test trial blocks (see Table 1). The finding of a significant triple interaction (S × P × T) indicated that the response curves for the 12 experimental groups were not parallel and that the relationship between the response measure and the two independent variables varied with the test trial block used in its determination. The significant Shock Level × Test Trial interaction (S × T) indicated, furthermore, that the persisting effects of shock intensity were not consistent throughout the test period. The test period was therefore decomposed into smaller segments before the Between *Ss* comparisons were evaluated.

TABLE 1

COMBINED ANALYSIS OF VARIANCE FOR DAYS 3 AND 4

Source	*df*	*F*
Between *Ss*	95	
Shock levels (S)	2	13.71**
CS-UCS pairings (P)	3	4.57**
S × P	6	.70
Error (b)	84	(.6950)[a]
Within *Ss*	864	
Test trials (T)	9	65.33**
S × T	18	10.04**
P × T	27	.96
S × P × T	54	1.47*
Error (w)	756	(.0406)[a]
Total	959	

[a] Error estimate.
* $P < .05$.
** $P < .01$.

TABLE 2

SEPARATE ANALYSES OF VARIANCE FOR DAYS 3 AND 4

Source	*df*	Day 3 *F*	Day 4 *F*
Between *Ss*	95		
Shock levels (S)	2	5.37**	20.11**
CS-UCS pairings (P)	3	4.71**	3.71*
S × P	6	.17	1.30
Error (b)	84	(.2789)[a]	(.5178)[a]
Within *Ss*	384		
Test trials (T)	4	153.26**	2.50*
S × T	8	3.86**	5.96**
P × T	12	1.09	1.12
S × P × T	24	.79	1.61*
Error (w)	336	(.0341)[a]	(.0319)[a]
Total	479		

[a] Error estimate.
* $P < .05$.
** $P < .01$.

Table 2 shows the results of separate analyses of the response measures obtained on Days 3 and 4. The triple interaction was significant on the second test day but not on the first; the S × T interaction reached the 1% level on both days. The form of the significant Shock Levels × Test Trial Block interaction is depicted in Fig. 1. The three curves contained in this figure were constructed by combining all levels of CS-UCS pairings for each shock intensity level: each curve therefore represents the

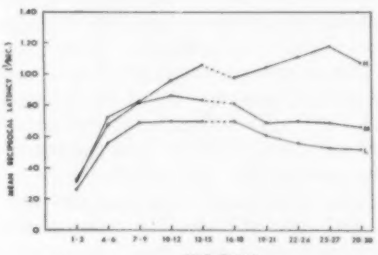

FIG. 1. Effect on performance of conditioning with high (H), medium (M), and low (L) shock intensities.

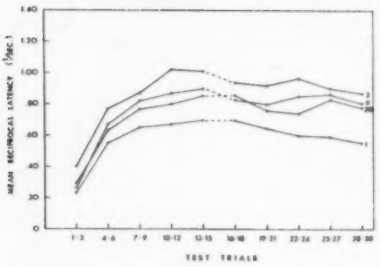

FIG. 2. Effect on performance of conditioning with 1, 3, 9, and 20 CS-UCS pairings.

means of successive response measures yielded by 32 *Ss* that were conditioned with the same intensity. Figure 1 shows that reciprocal latencies of high-shock groups were depressed on the early trials of the first test day, recovered on the later trials, and exhibited further growth during the second test day. The medium- and low-shock groups, on the other hand, reached maximum response strength by the end of Day 3 and declined thereafter. Figure 2, constructed in a similar manner, shows the effect of variations in the number of CS-UCS pairings on the response measure at successive portions of the test period. For this figure, all levels of shock intensity were combined for each CS-UCS pairing level; each curve represents repeated measurement of 24 *Ss*. The nonsignificant interaction between CS-UCS Pairings and Test Trial Blocks (P × T in Table 2) indicates that the hypothesis of parallel curves cannot be rejected for Fig. 2. The effect of this independent variable, therefore, was consistent throughout the test period.

The results of analyzing the test period into still smaller components are presented in Tables 3 and 4. Each table contains summaries of analyses of variance computed using data from two adjacent test trial blocks. Both tables show that all significant Within *Ss* interactions were thereby eliminated, thus permitting evaluation of Between *Ss* effects for the last, and next to last, block of six test trials on each test day. Figures 3A-D illustrate the relationship between the response measure and the independent variables at successive portions of the test period. The shock intensity effect is depicted as a parameter of a family of three curves which describe the response measure as functions of the number of CS-UCS pairings. Each point represents 8 *Ss* that were

TABLE 3

ANALYSES OF VARIANCE FOR TRIALS 4–9 AND 10–15 (DAY 3)

Source	*df*	Trials 4–9 *F*	Trials 10–15 *F*
Between *Ss*	95		
Shock levels (S)	2	2.86	6.99**
CS-UCS pairings (P)	3	2.95*	4.20**
S × P	6	.36	.23
Error (b)	84	(.1377)[a]	(.2167)[a]
Within *Ss*	96		
Test trials (T)	1	49.43**	.83
S × T	2	.25	2.54
P × T	3	.56	.34
S × P × T	6	1.61	.70
Error (w)	84	(.0144)[a]	(.0306)[a]
Total	191		

[a] Error estimate.
* $P < .05$.
** $P < .01$.

TABLE 4

ANALYSES OF VARIANCE FOR TRIALS 19–24 AND 25–30 (DAY 4)

Source	*df*	Trials 19–24 *F*	Trials 25–30 *F*
Between *Ss*	95		
Shock levels (S)	2	17.14**	25.24**
CS-UCS pairings (P)	3	3.60*	3.68*
S × P	6	.87	2.35*
Error (b)	84	(.2463)[a]	(.2469)[a]
Within *Ss*	96		
Test trials (T)	1	.13	5.43*
S × T	2	2.38	2.33
P × T	3	1.37	.08
S × P × T	6	.75	1.30
Error (w)	84	(.0216)[a]	(.0211)[a]
Total	191		

[a] Error estimate.
* $P < .05$.
** $P < .01$.

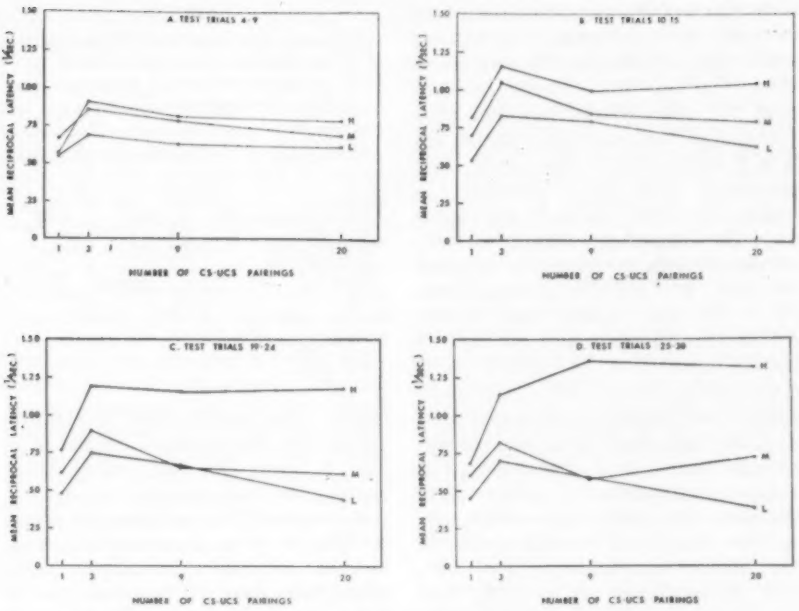

FIG. 3. Response strength as a joint function of number of CS-UCS pairings and shock intensity at successive portions of the test session.

conditioned with one of the 12 combinations of the independent variables. Comparison of Fig. 3A-D shows that shock intensity exerted greater influence with progressive test trials. This interpretation is supported, of course, by the significant S × T interaction previously described in Fig. 1 and reported in Tables 1 and 2. It receives further support from *F* ratios computed for the main effect of shock intensity at successive stages of the test period. These ratios did not reach significance until the end of Day 3 (Table 3 and Fig. 3B), but they remained significant, thereafter, well beyond the 1% level (Table 4 and Fig. 3C, 3D). The CS-UCS pairings variable, on the other hand, contributed significantly to Between *Ss* variance, beyond the 5% level, throughout the four test periods (Tables 3 and 4).

The predicted interaction between shock levels and number of CS-UCS pairings (S × P) was significant beyond the 5% level only at the end of the second test day (Table 4). Figure 3D illustrates the form of this interaction and should be compared with Fig. 3B which shows the relationships established at the end of the first test day. The nonsignificant S × P interaction (Table 3) obtained for Test Trials 10–15 indicates that the hypothesis of parallel curves cannot be rejected for Fig. 3B. The significant S × P interaction shown in Fig. 3D reflects further increments in response strengths of high shock *Ss*, particularly the 9- and 20-trial groups, and decrements in speeds of medium and low shock *Ss*.

A series of F ratios was computed in order to test several hypotheses concerning the shapes of the functions shown in Fig. 3D. These were not post hoc comparisons but had been planned in the event of a significant S × P interaction. The error term employed was the Between *Ss* residual for Test Trials 25–30 (Table 4). The hypotheses that the three curves coincide could not be rejected for the 1 CS-UCS pairing level ($F = .92$; $df = 2/84$), but it was rejected for the 3, 9, and 20 CS-UCS pairing levels ($F = 3.43, 13.20, 14.74$, respectively; $df = 2/84$). Comparisons of two curves at a time indicated that the high shock level was responsible for the S × P interaction: significant interaction F ratios were obtained for the high shock vs. medium shock and the high shock vs. low shock comparisons ($F = 3.09, 3.31$, respectively; $df = 3/84$), but not for the medium shock vs. low shock comparison ($F = .66$; $df = 3/84$). In addition, the hypothesis that the curves for the medium and low shock groups coincide could not be rejected ($F = 2.96$; $df = 1/84$).

Table 5 presents the results of a comparison between the control group and the experimental group that received 20 CS-UCS pairings with high shock. The hurdle-jump acquisition curves for these groups, contained in Fig. 4, show that the control group learned the test response on Day 3 but decreased its performance speed on Day 4. The experimental group, however, responded at a significantly faster rate than the control: the F ratio computed for the Between *Ss* effect was significant beyond the 5% level on both days (Table 5). The significant Within *Ss* interaction (Table 5) for Day 4 indicates that the two curves diverged on the second test day.

TABLE 5

ANALYSES OF VARIANCE FOR COMPARISON OF CONTROL GROUP WITH HIGH SHOCK, 20 CS-UCS PAIRING EXPERIMENTAL GROUP

Source	*df*	Day 3 *F*	Day 4 *F*
Between *Ss*	15		
Control vs. Exp. (EC)	1	4.84*	5.20*
Error (b)	14	(.1666)[a]	(.6209)[a]
Within *Ss*	64		
Test trials (T)	4	20.87**	.84
EC × T	4	2.30	4.93**
Error (w)	56	(.0490)[a]	(.0349)[a]
Total	79		

[a] Error estimate
* $P < .05$.
** $P < .01$.

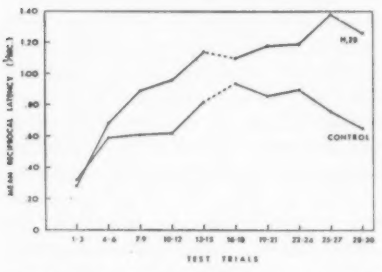

FIG. 4. Performance curves for the high shock, 20 CS-UCS pairings, experimental group and for the control group. $N = 8$ for each curve.

DISCUSSION

The major finding of this experiment is that variations in shock intensity and number of CS-UCS pairings introduced during aversive conditioning interact in their effect upon subsequent acquisition and performance of a hurdle-jump response only during the later stages of the test period. This result partly confirms the prediction, derived from a review of apparently contradictory studies of the conditioning trial parameter, regarding the importance of shock intensity as a potentially confounding variable in such determinations. The heterogeneity of the obtained interaction does not, however, correspond with the prediction. The particular form

exhibited by the interaction is, moreover, contrary to the expected one.

An examination of Kalish's (1954) data had led to the conclusion that the relationship between conditioning trials and maximum test performance could best be described, in that study, by a positively accelerated function. The notion that still higher shock levels would require fewer conditioning trials to attain a performance limit, thereby yielding a negatively accelerated function, had been proposed by Kalish (1954) and Mathers (1957). In the present experiment, the form of the predicted interaction was expected to reflect some combination of such curves. The obtained interaction (Fig. 3D) does indeed contain a negatively accelerated function at the high shock level, but the curves for medium and low shock groups are decidedly irregular in shape. Comparison of these functions with the curves shown in Fig. 3B, where the interaction is not significant, reveals that performance limits were reached at different stages of the test period by the three shock groups.

The mechanisms underlying the interaction are not immediately obvious, but it may represent nothing more than an artifact resulting from crude measurement of a complex test response. Long hurdle-jump latencies occurred not only when *Ss* failed to maintain their orientation toward the door, or were slow in crossing the barrier, but also when CS onset elicited behavior incompatible with the test response. Such competing responses were generally characterized by immobility, or else by disoriented, nongoal directed activity.

The most plausible interpretation of the interaction assumes that competing responses were conditioned during the shock session and that, varying in strength directly with shock intensity and number of CS-UCS pairings, they exerted the greatest influence among the high shock and the 9- and 20-trial groups. The differential effects of high shock were therefore obscured during early portions of the test session by crouching or by excessive activity. These interfering responses were elicited by CS onset, but they were not reinforced and thus gradually extinguished as test trials progressed. At the same time, hurdle-jump response tendencies were reinforced and continued to increase in strength. For the high shock groups, the net result of these opposing tendencies was the elevation of the entire gradient at the end of the test session, with a particularly marked increase exhibited in speeds of the 9- and 20-trial groups (Fig. 3D). The curves for medium and low shock groups, on the other hand, were distorted throughout the test period by competing responses of lesser magnitude that also extinguished. Since hurdle-jump tendencies reached maximum strength at an earlier portion of the test session for these groups, and then extinguished, the net effect was a marked decrease in speeds of the 9- and 20-trial groups at the end of the test session.

Another complicating factor that must be considered is the possible contribution of background cues to performance of the test response. Animals in the control group learned to cross the barrier, but they exhibited behavior qualitatively different from behavior of *Ss* in the comparable experimental group. Control animals did not learn to crouch: the hurdle-jump response occurred on the first test trial for all *Ss*. Long latencies on early test trials were, instead, accompanied by a high degree of disoriented activity. The control group was initially faster than the experimental, but the test response was less stable and showed some decrement on the second test day (Fig. 4). The learning curve for the control group may represent either an operant response level, a demonstration of backward conditioning, or the effect of conditioning to shock-compartment stimuli other than buzzers and lights. Since the last interpretation seems the most acceptable, it is probable that background cues served to enhance hurdle-jump tendencies during some portions of the test session, at least for the high-shock, 20-trial experimental group and perhaps for all experimental *Ss*.

The addition of these cues to the motivational complex could have raised the drive levels of all *S*s and could have further distorted the underlying relationships.

Acquired drive strength, in the present experiment, was not related to the response measure in any simple manner. Similar difficulties have been reported with this conditioning and test procedure (Brown & Jacobs, 1949) as well as with methods depending upon consummatory response inhibition (Libby, 1951) and escape-from-shock training (Dinsmoor & Campbell, 1956a). Kalish's (1954) study is one of the few that successfully minimized the influences of competing responses, other than certain disorienting tendencies, on early test trials. A major difference between his method and the present one may be his inclusion of an adaptation session between conditioning and test days. Although this additional session served a special purpose in his study, it may have been responsible, as well, for reduction in strength of crouching tendencies and for elimination of background cues before the initial test trial.

Future research in this area will encounter similar difficulties of interpretation unless some reliable method can be discovered that measures the incidence and strength of competing responses acquired during aversive conditioning or else eliminates completely the influence of such tendencies.

Summary

A parametric study of shock intensity and number of CS-UCS pairings, investigating a broad range of both variables, was conducted with an acquired drive procedure utilizing shocks of fixed, short duration during conditioning. The two variables were found to interact in their effect on subsequent performance of a hurdle jump response that was motivated by CS onset and reinforced by CS cessation; the interaction was significant, however, only at the end of the hurdle-jump session. The results were discussed in terms of possible uncontrolled factors contributed by the development of competing response tendencies.

References

Brown, J. S., & Jacobs, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.

Brown, J. S., Kalish, H. I., & Farber, I. E. Conditioned fear as revealed by magnitude of startle response to an auditory stimulus. *J. exp. Psychol.*, 1951, **41**, 317–328.

Dinsmoor, J. A., & Campbell, S. L. Escape-from-shock training following exposure to inescapable shock. *Psychol. Rep.*, 1956, **2**, 43–49. (a)

Dinsmoor, J. A., & Campbell, S. L. Level of current and time between sessions as factors in "adaptation" to shock. *Psychol. Rep.*, 1956, **2**, 441–444. (b)

Kalish, H. I. Strength of fear as a function of the number of acquisition and extinction trials. *J. exp. Psychol.*, 1954, **47**, 1–10.

Libby, A. Two variables in the acquisition of depressant properties by a stimulus. *J. exp. Psychol.*, 1951, **42**, 100–107.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Mathers, B. L. The effect of certain parameters on the acquisition of fear. *J. comp. physiol. Psychol.*, 1957, **50**, 329–333.

(Early publication received July 8, 1960)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 359–364

INSTRUMENTAL LICKING BEHAVIOR AS A FUNCTION OF SCHEDULE, VOLUME, AND CONCENTRATION OF A SACCHARINE REINFORCER [1]

STEWART H. HULSE, HARRY L. SNYDER, AND W. EDWARD BACON

Johns Hopkins University

This paper describes some data that were obtained when tongue licking in rats was instrumentally conditioned using two fixed-ratio schedules, two volumes, and three concentrations of a saccharine reinforcer. The system that was used to do this involved an electronic relay which automatically operated a pump each time *S* licked on a drinking tube. Consummation of fluid from the tube was thus made contingent upon the occurrence of the licking response itself.

The above approach bears upon recent work which has sought to determine the role that various factors correlated with ingestion may play in the more general process of reinforcement. Thus, Guttman (1953, 1954) and Hutt (1954) have shown, for example, that rate of lever-pressing is correlated with the concentration and quality of the agent used as a reinforcer. Collier and Siskel (1959), generalizing from similar data, have very cogently suggested that the reinforcing effect of a substance such as sucrose is an increasing function of its intensity and area of taste stimulation, but a decreasing function of its postingestive concentration. Because a dipper-delivered drop of fluid was used as the reinforcer in these experiments, however, they do not provide direct information concerning factors such as rate of ingestion and number of consummatory responses which have also been implicated as parameters of reinforcement (Kling, 1956). The present experiment was designed to examine such factors when schedule, volume, and concentration of reinforcement were systematically varied, and when a response which is intimately involved in the process of fluid ingestion was used.

[1] This research was supported in part by National Science Foundation Research Grant G8712 to the senior author. The authors wish to especially acknowledge the work and contributions of Leon S. Otis.

METHOD

Subjects

The *S*s were 48 experimentally naive male albino rats of the Wistar strain. They were approximately 80 days of age when received from Albino Farms, Red Bank, New Jersey.

Apparatus

The apparatus consisted of a 6.5-in. × 10-in. × 7-in. wooden box painted flat black with a Plexiglas lid and a metal floor. This box was placed inside a sound-resistant aluminum refrigerator box. A fan and two 3-w. bulbs located in the lid of the aluminum box provided, respectively, a masking noise and suitable light for *S*. A 2-in. × 3-in. opening was cut in one end wall of the inner wooden box. This opening was covered by a piece of ¼-in. Plexiglas which had a 1-cm × 2-cm. vertical slot cut in its center. A drinking tube was mounted behind the slot as described below.

The liquid reinforcement system used in this experiment has been described in detail elsewhere (Hulse, 1960). Briefly, it included a drinking tube and an electronic relay which operated each time *S*'s tongue touched a metal contact on the drinking tube. The output from the electronic relay operated a programming circuit. Liquid reinforcements were delivered by means of an infusion pump operated by the programmer. The pump consisted of a hypodermic syringe

which was driven known amounts by a Ledex rotary solenoid operating through a gear and lead-screw system. Each time a reinforcement was delivered, a solenoid-driven lever, also operated by the programmer, slightly squeezed but did not close the pressure tubing connecting the pump to the drinking tube. After a reinforced lick, the tubing was allowed to return to its normal shape. The resulting slight negative pressure drew the fluid column about 1 mm. inside the tip of the drinking tube, thus assuring that *S*'s tongue would not come in contact with the fluid except on reinforced licks.

Two drinking tubes were used with the system in this experiment, a $\frac{1}{8}$-in. diameter brass tube, and a $\frac{1}{4}$-in. diameter Plexiglas tube. Both tubes had 2-mm. fluid holes in their centers. In order to provide an electrical contact for the electronic relay, a piece of $\frac{1}{16}$-in. brass rod was pressed-fit through a second hole in the Plexiglas tube such that its end was flush with the tube's tip and located $\frac{1}{16}$ in. below the fluid hole.

Procedure

Experimental design.—Two fixed-ratio schedules of reinforcement (5.5:1 and 1:1),[2] two volumes of reinforcement on reinforced licks (.0023 cc. and .0074 cc.), and three concentrations of saccharine (0.5 gm., 1.5 gm., and 3.5 gm. added to 1 liter of tap water) were combined in a factorial design. Four *Ss* were run in each of the 12 conditions required by the design.

For convenience, the groups were assigned labels which summarized the treatment they received, e.g., 1.5-L-C. The initial number (0.5, 1.5, or 3.5) stands for the appropriate concentration of saccharine; the middle letter (L or S) stands for large or small volumes of reinforcement, and the final letter (P or C) stands for partial or continuous schedules of reinforcement.

Deprivation schedule and taming.—Upon receipt from the breeder, *Ss* were placed in colony cages and maintained on ad lib. Purina lab chow and water (obtainable from a drinking pan) for 20 days. At the end of this period, *Ss* were transferred to individual cages and placed on a daily diet of 7.5 gm. of ground Purina chow mixed with 10 cc. of tap water. This daily ration of wet mash was the only food and water *Ss* received throughout the remainder of the experiment, except for some extra water made available during taming as described below. The wet mash was given such that *Ss* were 22–23 hr. food and water deprived at the time the experimental procedure was started each day.

All *Ss* were tamed for 13 days. On Days 1–12, *Ss* were placed in groups of 4 or 5 in a large wooden box, where they were allowed 10 min. to explore and drink from a water pan. On Day 13, the procedure was the same except that the water pan was not available.

Acquisition.—Following taming, all *Ss* were given 12 days of acquisition of the instrumental licking response. On Day 1, *Ss* were allowed to lick from the brass tube, which extended $\frac{1}{4}$ in. into the box through the plastic shield, until each had made 200 licks on a 1:1 fixed ratio of reinforcement or until 20 min. had elapsed, whichever occurred first. The variables of concentration of saccharine and volume of reinforcement were introduced immediately. On Day 2, the procedure was identical except that the brass tube was withdrawn so that its tip was flush with the plastic shield. On Days 3 and 4, all *Ss* made 500 licks on the Plexiglas tube which was centered $\frac{1}{16}$ in. behind the plastic shield. The 5.5:1 fixed ratio of reinforcement was introduced for the appropriate *Ss*. On Days 5–12, the procedure was the same, except that the *Ss* made 700 licks.

Extinction.—Two days of extinction were given following acquisition. The procedure during extinction was exactly the same as during the last 8 days of acquisition, except that no fluid was placed in the hypodermic syringe of the infusion pump, and *Ss* thus licked on a dry tube. All *Ss* were extinguished to a 2-min. nonresponse criterion on each extinction day.

Measures of Performance

The time to the nearest second required for each *S* to emit the 700 licks on each of the last five days of acquisition was recorded as a measure of acquisition performance. These times were transformed to reciprocals and multiplied by 1000. The total number of responses emitted by each *S* on each of the two extinction days was recorded as a measure of resistance to extinction. Means of the reciprocals and of the response scores

[2] With the programing equipment available at the time the experiment was begun, it was possible to obtain a schedule of partial reinforcement in which reinforcement was delivered with the last lick in alternating blocks of 5 and 6 licks. The 5.5:1 ratio thus represents a mean value and is, of course, a "fixed" ratio in a somewhat limited sense.

were determined for each *S* for each day for purposes of the statistical analyses.

As an additional measure, mean volume of fluid per lick was calculated for the last day of acquisition. This measure was obtained by dividing total amount of fluid consumed (one of four fixed values determined by the possible combinations of the two schedules and two total volumes of reinforcement) by 700, the number of licks *S*s made.

Results

The analysis of variance which was run on the reciprocal drinking times for acquisition (Table 1) showed that the main effects of concentration of saccharine and Schedule of reinforcement are significant (*P*s < .05), and that the interaction between Volume and Schedule of reinforcement is also significant ($P < .01$). Neither the main effect of Days nor the interaction of Days with any of the other variables is significant (*P*s > .05); thus, there is no evidence that the mean performance of the groups was changing over the last five days of acquisition. None of the other main effects or interactions is significant (*P*s > .05).

TABLE 1

Analyses of Variance of Reciprocal Drinking Times × 1000 for the Last 5 Days of Acquisition and of Number of Licks for the 2 Days of Extinction

Source	Acquisition		Extinction	
	df	*MS*	*df*	*MS*
Days (D)	4	0.88	1	608,653.5**
Ss	47		47	
Concentration (C)	2	14.65*	2	1,206.5
Schedule (S)	1	10.40	1	273,280.0**
Volume (V)	1	25.00*	1	240,200.0**
C × V	2	0.25	2	13,215.8
C × S	2	3.05	2	1,830.3
V × S	1	119.10**	1	106,001.1**
C × V × S	2	0.45	2	4,045.2
Error (within)	36	3.37	36	10,154.4
Ss × Days	188	1.10	47	8,195.9
C × D	8	0.69	2	4,728.5
S × D	4	0.03	1	65,730.7**
V × D	4	0.07	1	57,428.2**
C × V × D	8	0.54	2	4,158.1
C × S × D	8	0.46	1	436.6
V × S × D	4	0.50	1	9,334.1
C × V × S × D	8	0.30	2	4,899.2
Error (within × Days)	144	0.97	36	6,229.7

* $P < .05$.
** $P < .01$.

TABLE 2

Mean Reciprocal Drinking Times × 1000 for the Last 5 Days of Acquisition and Mean Total Licks for the 2 Days of Extinction

Group	Acquisition (1000/time)	Extinction (total licks)
0.5-S-C	4.84	76.6
1.5-S-C	6.00	109.4
3.5-S-C	5.48	77.6
0.5-S-P	4.20	292.8
1.5-S-P	4.48	265.8
3.5-S-P	4.67	224.6
0.5-L-C	2.63	36.8
1.5-L-C	3.85	50.8
3.5-L-C	3.69	75.4
0.5-L-P	4.87	68.1
1.5-L-P	5.38	97.4
3.5-L-P	5.39	118.1

Note.—As an aid to the computation of *t* tests among the group means appearing in the table, the appropriate error variances for training and extinction are 3.37 and 10,154.4, ($df = 36$), respectively.

As shown by Table 2, the groups which received the 0.5 gm/liter concentration took longer to emit 700 licks than the groups which received the 1.5 gm/l concentration ($t = 3.85$, $df = 36$, $P < .01$) or the 3.5 gm/l concentration ($t = 3.27$, $df = 36$, $P < .01$). The latter two groups do not differ significantly from each other, although Table 2 suggests a tendency for the 3.5 gm/l concentration to produce somewhat slower responding than the 1.5 gm/l concentration under continuous-reinforcement conditions, but somewhat faster responding under partial-reinforcement conditions.

The significant interaction between Volume and Schedule of reinforcement is diagrammed in Fig. 1. Partial reinforcement resulted in faster drinking times than continuous reinforcement with a large volume of fluid on reinforced licks, but slower drinking times with a small volume of fluid on reinforced licks. By *t* tests, the

differences among the groups appearing in Fig. 1 are all significant ($Ps < .01$), with the exception of the difference between the S-C and the L-P groups which is not significant.

The analysis of variance run on the mean number of responses emitted during extinction (Table 1) shows that the main effects of Schedule of reinforcement and Volume of reinforcement, together with the interaction between these two variables, are significant ($Ps < .01$). Table 2 shows that, in general, partial reinforcement produced more licks than continuous reinforcement, and large volumes of reinforcement produced *fewer* licks than small volumes of reinforcement. However, Fig. 2, which diagrams the significant interaction between these two variables, shows that the partial vs. continuous reinforcement difference was much larger if small rather than large volumes of reinforcement had been used during training. By t tests, the differences among the groups appearing in Fig. 2 are all significant beyond the .01 level with the exception of the difference between the L-P group and the S-C group which is significant beyond the .05 level.

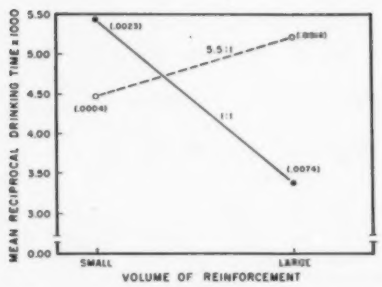

FIG. 1. Acquisition drinking time as a function of schedule and volume of reinforcement. (The numbers in parenthesis show mean volume of fluid per lick. The data points have been connected with straight lines, but no assumptions are intended concerning the actual shapes of the functions.)

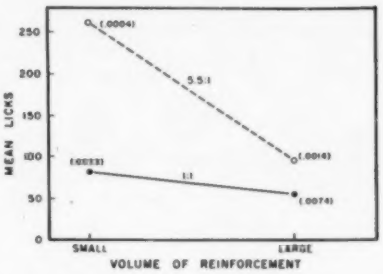

FIG. 2. Licks to extinction criterion as a function of schedule and volume of reinforcement. (The numbers in parenthesis show mean volume of fluid per lick. The data points have been connected with straight lines, but no assumptions are intended concerning the actual shapes of the functions.)

Figure 2 also shows that a perfect inverse ordering holds between mean volume per lick during training and the number of responses emitted during extinction.

Finally, Table 1 shows that the main effect of Days, and the Schedule × Days and Volume × Days interactions are significant ($Ps < .01$). In general, significantly fewer responses were emitted by all *Ss* on Day 2 than on Day 1 of extinction. In addition, the differences among the groups based on Volume and Schedule of reinforcement became smaller on Day 2 than on Day 1, but the rank order of the means of the groups did not change.

DISCUSSION

Mean volume of reinforcement per response, as jointly determined by the ratio of reinforced to nonreinforced responses and by the absolute volume of fluid per reinforcement, appears to be a significant parameter of licking behavior and, perhaps, of other types of instrumental behavior as well. The present acquisition data show that the

average rate[3] with which 700 licks were emitted on each of the last several days of acquisition was a nonmonotonic function of the mean volume of fluid obtained per lick. Average licking rate increased with increases in mean amount per lick with the exception that the L-C group, which received the largest mean amount, showed the slowest average licking rate. The mean-volume data are thus analogous to those data obtained in lever pressing situations where a nonmonotonic relationship has been found between rate of responding and the concentration of a sucrose solution used as a reinforcer (Collier & Siskel, 1959; Guttman, 1953). Thus, up to a point, rats assail the drinking tube or the lever as if increases in mean volume of fluid per lick or in concentration of sucrose were acting to increase the effectiveness of the reinforcer. Beyond some optimum point, however, a sufficiently large mean volume or a sufficiently high concentration may be reached such that fluid is coming in such large quantities with each lick that the animal has difficulty in accommodating it all, or because relatively long-term postingestive factors such as stomach loading or changes in taste sensitivity assume greater and greater importance with successive reinforcements. In either case, the animal adjusts by reducing its rate of lever pressing or its average rate of tongue licking. McCleary (1953) and Collier and Siskel (1959) have suggested that a system of the type outlined above holds for behavior based on different concentrations of sucrose; apparently, a similar system can easily account for the present data based on mean volume of reinforcement.

The concept of mean volume of reinforcement also accounts for the extinction data. First of all, the extinction data show, classically, that partial reinforcement results in greater resistance to extinction than continuous reinforcement. The data also show quite conclusively that large absolute volumes of reinforcement produce *less* resistance to extinction than small absolute volumes of reinforcement, a fact which runs counter to classic Hullian theory (Hull, 1943; Zeaman, 1949), but a fact for which there is increasing experimental support (Armus, 1959; Hulse, 1958). Most significantly, however, the number of licks emitted by the several groups during extinction is a perfect decreasing function of the mean volume of fluid received per lick during acquisition.

Finally, the saccharine data are of interest since they clearly parallel data obtained with responses less directly involved in the reinforcement process. Thus, mean drinking time during training tended to decrease as the concentration of saccharine reinforcement increased. This is directly analogous to lever pressing data which show, over a certain range of concentration of sucrose, that rate of responding increases with increases in concentration. The present data are not quite so congruent with Collier and Siskel's (1959) observation that response rate continues to increase with continued increases in sucrose concentration as long as the interreinforcement interval is relatively long, but that the relationship between concentration and response rate becomes nonmonotonic if interreinforcement intervals are relatively short. There is a tendency in the present data (cf. Table 2), however, for the decrease in mean drinking time to be somewhat greater if fluid ingestion rate is reduced through the use of partial vs. continuous reinforcement and small vs. large absolute volumes of reinforcement. Quite probably, a nonmonotonic relationship could be demonstrated more clearly by appropriately increasing concentration, rate of ingestion, and duration of the experimental session, i.e., by increasing the relative influence of long-term postingestive factors.

[3] Discussion in this paper is confined to average licking rate, a measure which is derived from the average time required to emit 700 licks. Early evidence indicates that momentary licking rate, the rate within a particular period of licking of a few seconds duration, may be correlated in an entirely different fashion with some of the variables under discussion here.

Summary

A new technique was employed to study instrumental licking behavior in rats as a function of three concentrations of saccharine, two fixed-ratios of reinforcement, and two volumes of fluid on reinforced licks. The time required for each of 48 rats to emit 700 licks on each of the last 5 days of a 12-day acquisition period was recorded as an index of training performance. The total number of responses emitted to a 2-min. nonresponse criterion on each of two successive days of extinction was recorded as an index of extinction performance.

The acquisition data show that mean drinking time systematically decreased with increases in concentration of saccharine, although there was a tendency for this relationship to become nonmonotonic. A definite nonmonotonic relationship was found between mean drinking time and mean volume of reinforcement per lick, as jointly determined by volume of fluid per reinforcement and ratio of reinforced to nonreinforced licks. The extinction data show, among other things, that the number of licks emitted to criterion was a decreasing function of the mean volume of reinforcement used during training.

In discussing the data, special emphasis was placed on the promise of mean volume of fluid per response as an important parameter of licking behavior, and, perhaps of the more general process of reinforcement.

References

ARMUS, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, **58**, 61–63.

COLLIER, G., & SISKEL, M., JR. Performance as a joint function of amount of reinforcement and interreinforcement interval. *J. exp. Psychol.*, 1959, **57**, 115–120.

GUTTMAN, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as a reinforcing agent. *J. exp. Psychol.*, 1953, **46**, 213–224.

GUTTMAN, N. Equal-reinforcement values for sucrose and glucose solutions as compared with equal sweetness values. *J. exp. Psychol.*, 1954, **47**, 358–361.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HULSE, S. H., JR. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

HULSE, S. H., JR. A precision liquid feeding system controlled by licking behavior. *J. exp. Anal. Behav.*, 1960, **3**, 1–3.

HUTT, P. J. Rate of bar pressing as a function of quality and quantity of food reward. *J. comp. physiol. Psychol.*, 1954, **47**, 235–239.

KLING, J. W. Speed of running as a function of goal-box behavior. *J. comp. physiol. Psychol.*, 1956, **49**, 474–476.

MCCLEARY, R. A. Taste and post-ingestive factors in specific-hunger behavior. *J. comp. physiol. Psychol.*, 1953, **46**, 411–421.

ZEAMAN, D. Response latency as a function of amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

(Received November 6, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 365–370

DRIVE, VERBAL PERFORMANCE, AND MUSCLE ACTION POTENTIAL

JOSEPH B. SIDOWSKI [1]

San Diego State College

AND

ROBERT G. EASON

United States Navy Electronics Laboratory

The present study investigates the relationship between drive, muscle action potentials, and performance in a serial rote learning task.

The acceptance of drive as an energizer, and of the multiplicative relationship between drive and habit strength, commits one to a number of implications about drive interaction and learning. One implication is that drive favors whatever response has the greatest habit strength, at the expense of all other responses; a second implication is that several drive states should have an additive effect on performance.

In considering the first implication, studies using high- and low-anxious drive levels (Taylor, 1953) have indicated that increased drive may facilitate simple learning and interfere with more complex learning (Montague, 1953; Spence, 1958; Taylor, 1956). Spence (1958) interprets these results within the framework of a theory of emotionally based drives.

In studying the implications of additive drive effect, Meyer and Noble (1958) attempted to show that Low-Anxious *Ss* make fewer errors in verbal maze learning while working under induced muscular tension, whereas High-Anxious *Ss* perform better without tension. Meyer (1953) explains the influence of anxiety and induced tension within the framework of a theory of simultaneous interaction of neural responses. In this theory, drive is considered a response and its capacity to alter the properties of other simultaneous responses stems from their convergence upon a final common path within the efferent mechanism. In this instance, the hand gripping a dynamometer, clearly a response has a property of a drive which should summate with conventional drives, such as anxiety, to alter performance (Meyer & Noble, 1958). Meyer also proposes that a high level of muscle activity makes learning more rapid, at least early in learning, and that the optimal amount of tension (either naturally occurring or induced) is progressively less as learning proceeds.

Induced muscular tension effects have been studied in a number of verbal learning experiments (Meyer, 1953). There is a scarcity, however, of information concerning the relationships between normally occurring muscle action potentials (MAPs), induced muscular tension and anxiety in verbal learning tasks (Berry & Davis, 1958; Malmo, 1958). In the present experiment, therefore, MAPs were recorded from High- and Low-Anxious *Ss* during learning under conditions of induced muscular tension and incentive.

[1] The authors gratefully acknowledge the helpful aid of Norman H. Anderson in the preparation of the final manuscript.

METHOD

Subjects.—Sixteen introductory psychology students at San Diego State College served as *Ss*, for which they received class points. Eight of the *Ss* were randomly selected from a high-anxious (highest 15%) and eight from a low-anxious (lowest 15%) population as determined by the Taylor anxiety scale (1953). The Low-Anxious *Ss* had a mean score of 6.6 (range 4–9) and the High-Anxious had a mean score of 27.0 (range 24–32). The only restriction on selection was one of equal sex representation; four men and four women in each group.

Task.—Each *S*, acting as his own control, learned a list of nonsense syllables while serving under each of the following conditions: Control (merely learn a list); Dynamometer (induced muscular tension while learning); Incentive (reward for fastest learner); and Incentive plus Dynamometer (reward for fastest learner under induced muscular tension). Muscle action potentials (MAP) were recorded from biceps and frontalis muscles during each condition.

Design.—Two independently drawn 4 × 4 greco-latin squares were used for each anxiety group. Latin letters represented conditions while Greek letters represented lists.

Apparatus and procedure.—During learning sessions *S* sat in a shielded room to permit MAP recording. Muscular tension was induced and recorded with a strain-gauge wrist dynamometer attached to the right side of a heavy arm chair. During the experimental session a padded stirrup was positioned on the volar side of the wrist. The *S*'s elbow was cupped in a padded L-shaped wooden box which provided a stationary fulcrum for elbow flexion. Flexion of the elbow produced pressure against the wrist stirrup, the force of which was relayed by the strain-gauge to an amplifier and dynograph recorder. The flexion resulted in isometric contraction of biceps, as well as certain muscle groups in the forearm. Two commercial silver disc electrodes, 8 mm. in diameter, were attached to the skin, 2 in. apart, over the biceps muscle for measuring localized potentials associated with elbow flexion. An additional set of electrodes was attached to the skin over the frontalis muscle and constituted a measure of general muscle tension. The skin resistance at the point of electrode contact was routinely checked for each *S* before each condition and was kept below 5,000 ohms. The associated MAP changes were led to pulse frequency type electronic integrators, the outputs of which were channeled into the dynograph of an Offner Type D EEG. Unintegrated MAP activity was simultaneously recorded with two channels of the EEG, and was monitored for electronic artifacts. A specific description of the apparatus and method is given elsewhere (Eason, 1959).

Each *S* was seated in front of a glass window through which he could observe a red signal light and a meter indicating the amount of force being applied to the dynamometer. The light circuit was arranged so that a prescribed level of force resulted in the light going off. Deviations from the required force by plus or minus 1 lb. caused the light to go on. The amount of force each *S* was required to apply during the Dynamometer conditions was defined as one-fourth of that which he was capable of exerting after 30 sec. of continuous maximal contraction.

Five lists of nonsense syllables, one practice and four experimental, taken from Glaze (1958) and of high association value (87–100%) were used. A different consonant started each of the syllables in a list with the vowels being used about equally often. This was done in order to reduce intralist similarity. Each list consisted of 10 nonsense syllables. Syllables were presented auditorily by a tape recorder at a 4-sec. rate with an 8-sec. rest interval between trials. Since the first syllable served as an anticipatory cue, *S* learned only nine items. Subjects learned each list to a criterion of one correct anticipation. A 15-min. rest interval followed each condition.

Under all conditions, *Ss* were told to learn a list as quickly as possible. When serving under either of the two Dynamometer conditions, *Ss* were also told to exert the required amount of force on the dynamometer and thereby to keep the light off throughout a trial. Subjects relaxed during the 8-sec. intertrial intervals. During the Control condition, *Ss* were told merely to learn the list and to keep their eyes on the light at all times. Instruction for the Incentive condition was identical with the Control except that *Ss* were told that the fastest learner would receive $1.00 plus an additional class point and the second fastest would receive an additional class point. The Dynamometer plus Incentive condition combined the use of the wrist dynamometer with the incentives. The two Incentive conditions, therefore, afforded each *S* the opportunity to earn $2.00 and two extra class points. At the conclusion of the sessions, *Ss* were questioned as to the value of the money and class points to determine whether they had

incentive value. All *Ss* expressed interest in winning both the money and class points.

Results

Learning data.—Anticipatory errors and trials to criterion were used as measures of performance. An inspection of the number of errors per syllable showed a classical bowed serial effect for all but the Dynamometer condition. For the latter, the number of errors was greatest for the fourth rather than the sixth syllable in the list for both anxiety groups. A statistical analysis of the total number of errors showed no significant effects for either the anxiety or learning variables. The data did, however, show a much larger number of errors for the dynamometer conditions. An additional analysis, therefore, was conducted to determine whether the combined induced muscular tension conditions (Dynamometer and Dynamometer + Incentive) differed significantly from the combined nondynamometer (Control and Incentive) conditions. Analyses of variance showed a significantly larger number of errors for the combined dynamometer conditions, $F = 9.03$, $df = 1/14$, $P < .01$. No other statistically significant differences were found.

An analysis of variance of number of trials to the criterion of nine correct is present in Table 1, and shows a significant Conditions effect. Duncan's (1955) test for differences between ranked means showed the Dynamometer condition required significantly more trials than any other for the Low-Anxious group. High-Anxious *Ss* took significantly more trials under the Dynamometer, and Dynamometer plus Incentive conditions, than under either of the remaining conditions. A subsequent analysis of the significant Ordinal Position effect showed the last experimental session requiring significantly fewer trials to learn the list than the first session. No other significant differences, however, were found between any of the other experimental session comparisons.

TABLE 1

Analysis of Variance of Trials to Criterion of Nine Correct

Source	*df*	*MS*	*F*
Anxiety (A)	1	1.00	
Ss within Groups	14	79.60	
Ordinal position (OP)	3	35.00	2.99*
Lists (L)	3	17.00	1.45
Conditions (C)	3	38.30	3.27*
A × OP	3	10.33	
A × L	3	.67	
A × C	3	8.00	
Error	30	11.70	
Total	63		

* $P = .05$.

The dynamometer conditions were again combined for analysis of the trial measures. Analysis of variance revealed that the combined dynamometer conditions required more trials to learn the list than the nondynamometer conditions, $P < .01$.

Analyses of the data on trials to criteria of 1, 2, 3, 4, and 5 correct showed no significant differences between either the anxiety or learning variables.

Previous investigations have often used performance scores over the early trials (Montague, 1953; Spence, 1958). In order to get comparable results, an analysis was made on total correct anticipations over the first two trials. The mean score of 4.5 for Low-Anxious *Ss* was significantly greater than the mean score of 2.8 for High-Anxious *Ss*, $F = 4.73$, $df = 1/14$, $P < .05$.

MAP data.—MAP recordings from biceps were used for monitoring

purposes only and were not included in the statistical analyses. It may be noted, however, that the biceps MAPs showed no apparent variation either between conditions or over trials within conditions.

The frontalis potentials were integrated over each trial and expressed in terms of mean MAP level in microvolts. These trial measures were subjected to analysis.

An initial analysis was conducted to ascertain whether there were any significant changes in MAP over the first, middle, and last trials of a learning session. Table 2 shows the MAP output in microvolts as a function of trial position. It is apparent that the High-Anxious *Ss*' MAPs decreased over the three trial positions, while the Low-Anxious *Ss*' MAPs increased. The reliability of this differential trial trend is shown by the significant Groups × Trials interaction, $P < .05$. Table 3 shows the mean MAP level based on the three trial positions and plotted as a function of experimental conditions. With the dynamometer conditions, MAPs were higher for the High-Anxious group. Under the non-dynamometer conditions, the opposite relation between the two anxiety groups was found. Analysis of variance showed this Groups × Conditions interaction to be significant, $P < .01$. No other significant differences were found.

TABLE 2

MEAN MAP LEVELS (AVERAGED ACROSS THE FOUR CONDITIONS) FOR HIGH- AND LOW-ANXIOUS GROUPS AS A FUNCTION OF TRIAL POSITION

Anxiety	Trials		
	First	Middle	Last
Low	26.9	27.7	28.4
High	27.5	24.8	24.1

TABLE 3

MEAN MAP LEVELS (AVERAGED ACROSS THE FIRST, MIDDLE, AND LAST TRIALS) FOR HIGH- AND LOW-ANXIOUS GROUPS AS A FUNCTION OF CONDITIONS

Anxiety	Experimental Conditions			
	Control	Dyna-mometer	Incentive	Dyn. + Inc.
Low	29.0	25.8	29.3	25.4
High	24.0	27.3	22.8	28.3

High vs. low MAP.—Since Malmo (1958) has suggested that physiological measures be taken as indicants of motivational strength, it was decided beforehand to carry out an analysis on groups divided on the basis of MAP levels. Accordingly, *Ss* were divided into High and Low MAP groups on the basis of MAP measures summed over the four experimental conditions. Analysis of variance on number of trials to the criterion of nine correct showed no significant difference between High and Low MAP groups, and no significant Condition × MAP Group interaction.

An analysis on High vs. Low MAP groups was also carried out on the number of correct anticipations during the first two trials, and no significant differences were found.

Mediocre vs. extreme learners.—Several studies have indicated that superior and inferior performers showed higher levels of muscle activity than mediocre performers (Bartoshuk, 1955; Berry & Davis, 1958). In this analysis, therefore, the four fastest and four slowest learners were placed in the Extreme group, and the remaining eight *Ss* in the Mediocre

group. The analysis was carried out on MAP measures which were recorded during each of the four experimental conditions. There was no significant difference between the Mediocre and Extreme learners, although the mean MAP level was much higher for the extremes. However, a significant Conditions × Group interaction was found ($F = 2.87$, $df = 3/42$, $P < .05$). Analysis of the interaction showed that for the extreme learners, MAP measures were highest for the Control and Incentive conditions. For the Mediocre learners, MAPs were highest for the two dynamometer conditions.

Discussion

On the basis of Meyer's (1953) motor interaction theory of drive, we would expect that the High-Anxious *Ss* would do better early in learning, but that the Low-Anxious *Ss* would excel if practice continued to a high criterion. Furthermore, the addition of tension would be expected to facilitate performance for the Low-Anxious *Ss* (Meyer & Noble, 1958).

The results of the learning data are contrary to Meyer's theory in every respect: (*a*) Low-Anxious *Ss* are superior to High-Anxious *Ss* early in learning, (*b*) this superiority decays later in learning, (*c*) the use of the wrist dynamometer interferes with performance, and (*d*) does so equally for High- and Low-Anxious *Ss*; finally, (*e*) the predicted summation effect of drives is not substantiated. With respect to the summation effect, it is interesting to note that the results of Meyer and Noble (1958) fail to support their own predictions concerning the interaction of induced muscular tension, manifest anxiety, and practice.

MAP and learning.—Although Spence (1958) merely suggests the importance of physiological indices of emotionally based drives, Meyer's theory (1953) is specific in predicting the relationships between level of muscle activity, drive, and performance.

For Meyer, high MAPs should be facilitative early in learning but detrimental later in learning. The significant Trials × Anxiety interaction for MAP measures are consistent with this prediction. (The Low-Anxious group, which has better verbal performance scores early but not late in learning, significantly increases in MAP over trials.) Since we also have a significant Conditions × Anxiety interaction with MAP scores, the theory would further predict that the verbal response scores would be proportional to the MAPs. We would, therefore, also expect a significant conditions by anxiety interaction in verbal scores, at least early in learning. No such interaction was found.

Considering the same scores, Taylor (1956) and Spence (1958) might predict that induced tension would be facilitative late in learning since competing responses would be weak relative to the correct response. If Spence were to use MAP as a measure of drive, a facilitation would also be predicted later in learning for high MAP. Therefore, if the significant Conditions × Anxiety MAP interaction were consistent over the last several trials, additional support would be rendered to the latter prediction. An inspection of the last two trials shows the data to be consistent with this prediction.

The results of the High vs. Low MAP groups categorization are interesting in light of the suggestions made by Malmo (1957, 1958) concerning the use of physiological activity as the key to distinguishing high-drive from low-drive *Ss*. Within the framework of the present study, data analyzed from groups separated on the basis of manifest anxiety scores proved more significant than those evaluated on the basis of total MAP measurements. It should be noted, however, that the tension found in different muscles does not always relate equally well to task performance (Berry & Davis, 1958). It is quite possible that

other muscle groups, or types of physiological measures, would have proven more meaningful.

Summary

High- and Low-Anxious *Ss* were required to learn lists of nonsense syllables under four conditions: Control (merely learn list), Dynamometer (learn list under induced muscular tension), Incentive (offered reward for fastest learner), and Dynamometer + Incentive (combined induced tension and reward). Muscle action potentials were recorded from the biceps and frontalis muscles throughout the sessions.

The results of the learning scores indicated: (*a*) Low-Anxious *Ss* superior to High-Anxious *Ss* but only over the early trials, and (*b*) Dynamometer pressure resulted in poorer learning for both anxiety groups. The muscle action potentials showed: (*a*) High-Anxious *Ss* gave higher MAPs than Low-Anxious *Ss* when the dynamometer was used, and with the opposite relation between Anxiety groups during the Control and Incentive conditions, and (*b*) MAPs decreased over trials for High-Anxious *Ss*, and increased for Low-Anxious *Ss*. Finally, (*a*) no difference was found in verbal performance between High and Low MAP groups, where the assignment of *Ss* to groups was done on the basis of total MAP scores, summed over trials and conditions, and (*b*) Mediocre learners showed higher MAPs than Extreme learners for the induced muscular tension conditions, while the opposite effect was obtained for the nondynamometer conditions.

The results were discussed within the framework of several theories.

References

Berry, R. N., & Davis, R. C. Muscle responses and their relation to rote learning. *J. exp. Psychol.*, 1958, **55**, 188–194.

Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

Eason, R. G. An electromyographic study of impairment and estimates of subjective effort associated with voluntary muscular contraction. *USN Electronics Lab. res. develpm. Rep.*, 1959, No. 898.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

Malmo, R. B. Anxiety and behavioral arousal. *Psychol. Rev.*, 1957, **64**, 276–278.

Malmo, R. B. Measurement of drive: An unsolved problem in psychology. M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958.

Meyer, D. R. The interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.

Meyer, D. R., & Noble, M. E. Summation of manifest anxiety and muscular tension. *J. exp. Psychol.*, 1958, **55**, 599–602.

Montague, E. K. The role of anxiety in serial learning. *J. exp. Psychol.*, 1953, **45**, 91–96.

Spence, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Taylor, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **43**, 303–320.

(Received November 9, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 371–375

EFFECTS OF INTERMITTENT REINFORCEMENT OF AN IRRELEVANT DIMENSION AND TASK COMPLEXITY UPON CONCEPT IDENTIFICATION [1]

LYLE E. BOURNE, JR. AND ROBERT C. HAYGOOD

University of Utah

The purpose of the present study was to evaluate the effect of intermittent reinforcement (IR) of an irrelevant dimension on performance in a two-choice concept identification task and on transfer to a second problem wherein this dimension was relevant to solution. The results of this experiment, in which a wide range of IR values were investigated at three levels of task complexity, extend available knowledge about the role of this variable in concept learning.

Gormezano and Grant (1958) found that IR of an irrelevant dimension increased the total number of trials required to attain criterion in the four-choice Wisconsin Card Sorting Test (WCST). Furthermore, they found no significant transfer effects to a second stage of sorting wherein the intermittently reinforced dimension of the first stage was relevant to solution. However, two other sets of experiments, using two-choice tasks, suggest that transfer should vary as a function of IR of an irrelevant dimension. On one hand, several studies (e.g., Buss, 1956) have shown that nonreversal shifts produce negative transfer; i.e., *S* makes more errors in reaching solution of the second problem than the first, when the relevant dimensions of the two problems are different. In such studies, the relevant dimensions of the first and second sorting stages were reinforced at a chance level (50%) in the second and first stages, respectively. Under these conditions, negative transfer would be expected since the shift required a rearrangement of the previously established S-R relationships (Gagné, Baker, & Foster, 1950; Porter & Duncan, 1953). Secondly, a study (Bourne & Haygood, 1959) aimed at determining the effects of redundant relevant dimensions on concept identification showed that the majority of *S*s were aware, when problem solution was reached, that more than one stimulus dimension could be used to classify the patterns correctly. Indeed, many *S*s reported exactly which dimensions could be used. Complete stimulus redundancy, as employed in the Bourne and Haygood experiment, is identical with the 100% R (regular reinforcement) condition of the present experiment. The earlier results suggest strongly that transfer to a second stage of sorting, in which only one of the previously redundant dimensions is relevant, would be positive.

On the basis of these findings, it would be expected that in a two-choice sorting task transfer should be negative with midrange values of IR of an irrelevant dimension, i.e., around 50%, but should change to positive between 50% and 100% reinforce-

[1] This research was supported in part by a research grant (M-1759) from the National Institute of Mental Health, United States Public Health Service, and was carried out during the second author's tenure of a Predoctoral Fellowship (MF-9829) from the National Institute of Mental Health.

ment. The present experiment was designed to explore this hypothesis by varying IR over six values between 50% and 100%.

METHOD

Task and procedure.—The task was essentially the same as that used in earlier concept identification studies (e.g., Bourne & Haygood, 1959). The *S* was required to sort or classify a series of 128 visually presented stimuli into two categories, which represented the two levels of the relevant dimension. The *S* indicated his choice of category for any stimulus by pressing one of two response buttons, neither of which bore a label. The stimuli consisted of geometric patterns which varied within binary dimensions such as color (red-green), form (square-triangle) and number (single-double figures). The total number of such dimensions, relevant plus irrelevant, determined the number of different stimuli within the series presented to *S*. At the outset, *S* was given detailed oral instructions describing the various dimensions, how they might be used in problem solution, the operation of pertinent apparatus, meaning of the reinforcing or feedback signal lamps, and the length of the problem. The presentation of stimuli, the control of delay (0 sec.) and duration (1 sec.) of information feedback or reinforcement, and of overall trial length (5 sec.), and the recording of responses were accomplished automatically with the concept identification apparatus described by Bourne and Haygood. The criterion of solution in Stage 1 of sorting was that no errors occur beyond stimulus pattern 48. All problems were self-paced.

Subjects.—The *S*s were 180 students from introductory psychology classes who were assigned by order of appearance to the 36 treatment combinations of the experiment. Thirty-one additional *S*s were dropped from the main analysis because of failure for various reasons to complete or to meet the criterion of Stage 1. These *S*s were distributed without bias over experimental conditions.

Design.—The $6 \times 3 \times 2$ repeated measures factorial design included six levels of IR of an irrelevant dimension (50%, 60%, 70%, 80%, 90%, and 100%), three levels of task complexity (1, 3, and 5 irrelevant stimulus dimensions in addition to the intermittently reinforced dimension), and two different problems, (*a*) color dimension relevant and number dimension intermittently reinforced during Stage 1 (first 64 patterns), shifting to number relevant and color irrelevant in Stage 2 (last 64 patterns), and (*b*) number relevant and color intermittently reinforced, shifting to color relevant and number irrelevant. The two stages were treated separately in the analysis of the results although experimentally they were continuous; *S* was not instructed about the shift. Since the tasks were two-choice, each irrelevant dimension, except that for which reinforcement varied in Stage 1, was reinforced 50% of the time. Henceforth, however, IR will be used only in reference to the irrelevant dimension for which this characteristic varied in the experimental design. It should be noted that in Stage 1 of the 100% R conditions there were two redundant relevant dimensions, either or both of which could be used by *S* to solve the problem. During Stage 2 one of these dimensions was reinforced only 50% of the time while the other remained relevant, as in Stage 1, without reversal of its levels.

RESULTS

Since instructions to *S* stressed accuracy, number of errors was the main dependent variable. However, it should be noted that variation in IR of an irrelevant dimension affected the probability of responses based on that dimension being correct. Using a dimension which was reinforced 50% of the time as a basis for responding, *S* had a probability of error equal to one-half; using a 90% reinforced dimension, he had an error probability of one-tenth or nine-tenths depending upon how he assigned the levels of that dimension to the response buttons. Such a systematic bias in the error measure suggests that the analysis of errors be supplemented by an analysis of total number of correct responses preceding the last error, an index of the number of reinforcements or reminders needed by *S* to confirm the correct sorting hypothesis (Gormezano & Grant, 1958; Grant & Curran, 1952). The error score for Stage 2 in this experiment was unaffected by such bias.

Figure 1 presents the mean number of errors made in the two sorting stages under each condition of IR and task complexity. Three trends are apparent: (*a*) total errors increase with number of irrelevant dimensions, (*b*) number of errors is inversely related to percentage IR of an irrelevant dimension, and (*c*) generally, total errors increase from Stage 1 to Stage 2, a trend which is even more marked when the fact that not all *S*s reached solution in Stage 2 is noted. The analysis of variance performed on these data is summarized in Table 1. Trend analyses, based on orthogonal polynomials, indicated that (*a*) the performance-irrelevant information function is linear, as in earlier results (e.g., Bourne, 1957), and (*b*) the IR function has significant linear and quadratic components. The quadratic component of this function is due primarily to the sharp drop in errors in Stage 2 at the level of 100% R. Statistical confirmation of this effect was provided by an orthogonal polynomial analysis of the IR×Stages interaction, in which only the quadratic term was significant. Problems represented a significant source of variance due to the fact that the number dimension was more difficult to use in sorting than color. Similarly, *S*s committed more errors when shifting from color relevant than from number relevant. The significance of the Irrelevant Dimensions×Stages interaction is due to a greater difference in difficulty between the levels of task complexity in Stage 2 than in Stage 1.

TABLE 1

ANALYSIS OF VARIANCE OF ERRORS

Source of Variance	*df*	*MS*	*F*
Intermittent reinforcement (IR)	5	271.58	7.55**
IR: Linear	1	910.94	25.34**
IR: Quadratic	1	143.35	3.99*
IR: Cubic	1	128.05	
IR: Quartic	1	120.54	
IR: Quintic	1	55.01	
Number of irrelevant dimensions (I)	2	729.15	20.28**
I: Linear	1	1372.82	38.18**
I: Quadratic	1	85.42	
Problems (P)	1	176.40	4.91*
IR × I	10	42.21	
IR × P	5	20.66	
I × P	2	2.05	
IR × I × P	10	58.40	
Ss/Groups	144	35.96	
Stage (S)	1	1504.70	50.04**
IR × S	5	72.48	2.41*
IR: linear × S	1	40.41	
IR: quadratic × S	1	151.57	5.04*
IR: cubic × S	1	70.41	
IR: quartic × S	1	20.59	
IR: quintic × S	1	79.45	
I × S	2	102.90	3.42*
P × S	1	199.50	6.63*
IR × I × S	10	29.73	
IR × P × S	5	35.42	
I × P × S	2	24.30	
IR × I × P × S	10	29.82	
Residual	144	30.07	

*$P < .05$.
**$P < .01$.

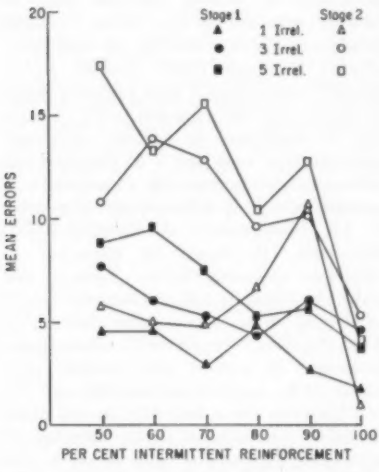

FIG. 1. Mean number of errors made in Stages 1 and 2 as a joint function of IR and task complexity. (Each plotted point represents the data from 10 *S*s.)

An analysis of variance was performed on total correct responses to last error in Stage 1. Only one main effect, number of irrelevant dimensions, was statistically significant ($F = 6.56$, $2/144$ *df*, $P < .01$). Under conditions of the present experiment, there is a general tend-

ency for number of correct responses to solution to *decrease* with increased IR, although the linear component of this function is not significant ($F = 3.88$, $1/144$ *df*, $.10 > P > .05$). For the six IR conditions, 50% through 100%, mean correct responses to last error were 9.0, 7.1, 7.2, 6.0, 7.4, and 4.6, respectively, and mean trials to last error were 16.2, 13.9, 12.4, 10.9, 12.3, and 7.9, respectively.

Discussion

Significantly more errors were made, generally, during Stage 2 than during Stage 1. This result agrees with those of Buss (1956) and Kendler and D'Amato (1955) which showed that the second concept in a nonreversal shift is more difficult to learn than the initial concept. However, the mean number of errors made by *S*s in the 100% R condition during Stage 2 was less than that made by 50% IR *S*s in Stage 1, regardless of the number of irrelevant dimensions. Since these groups performed under identical conditions save for the Stage 1 practice of the 100% R *S*s, the conclusion of Bourne and Haygood (1959) was supported. That is, since *S*s can generally describe all relevant dimensions when solution is attained, transfer from a problem containing two or more redundant relevant dimensions to a problem in which only one of these is relevant will be positive. The finding that performance in Stage 2 was significantly related to degree of IR in Stage 1 can be accounted for with conventional S-R theory. Negative transfer is predictable since, although *S* makes the same responses to the same set of stimuli in Stage 2, a rearrangement of the S-R connections is required. This condition, according to Gagné et al. (1950), produces maximum negative transfer. The decrease in negative transfer with increased IR is predictable also since reinforcement would increase presumably the habit loading for such an irrelevant dimension. Furthermore, it is conceivable that a decrease in negative transfer should occur because degree of IR determines, in part, the discriminability of the two stages of this experiment. When *S* is shifted from the 50% IR condition, the stimulus patterns remain the same in Stage 2 as they were in Stage 1. However, when the shift is made from the 100% condition, patterns not presented in Stage 1 occur in Stage 2. The frequency with which certain patterns occur differs between the two stages depending upon percentage of IR. This factor may contribute to decreased negative transfer by increasing the discriminability of the two stages for *S*. This argument is similar to that described by Jenkins and Stanley (1950) to account for an increase in trials to extinction due to partial reinforcement during acquisition.

Performance in Stage 1 was facilitated by IR of an irrelevant dimension as evidenced by a decrease in number of errors and number of correct responses to solution as IR increased. This result is in accord with the finding that redundant relevant information improves concept identification performance (Bourne & Haygood, 1959). In the case of neither measure is the trend a statistically significant function of task complexity although there appears to be greater improvement with higher degrees of irrelevant information. The result can be explained as follows. Concept identification requires *S* to discriminate relevant from irrelevant dimensions, a process which is affected by the level of IR (Gormezano & Grant, 1958). Increased IR would be expected to exert two opposing effects. First, it will make the relevant and the intermittently reinforced dimensions more distinctive than the other (irrelevant) dimensions. However, it should also inhibit performance by making the final discrimination between these two dimensions more difficult. An increase in the number of irrelevant dimensions will increase the facilitating effect of IR, while the inhibiting effect remains unchanged. Therefore, increased IR should result in

increased facilitation of performance as number of irrelevant dimensions increases.

Finally, it should be noted that, while the present experiment provides evidence of the importance of IR in concept identification, it was not designed to explore the differences between reversal and nonreversal shifts. For this purpose, additional studies wherein IR for the relevant dimension of Stage 1 is varied during Stage 2 are required.

Summary

An experiment was conducted to determine the effects of task complexity and of intermittent reinforcement (IR) of an irrelevant dimension upon the identification of concepts. In addition, the experiment was designed to measure transfer effects to a second task wherein the previously intermittently reinforced dimension was relevant and the previously relevant dimension was irrelevant to solution. The task required *S* to learn correct methods for classifying visually presented geometric stimuli. The 180 *Ss* served in a $6 \times 3 \times 2$ repeated measures factorial design with six levels of IR (50%, 60%, 70%, 80%, 90%, and 100%), three levels of task complexity (1, 3, and 5 irrelevant stimulus dimensions), and two different problems. Each *S* was required to serve in two successive stages of sorting.

The major findings were: (*a*) As IR increased in Stage 1, performance improved. (*b*) Generally, negative transfer from Stage 1 to Stage 2 was found with the amount of such transfer becoming progressively less with increased IR. The 100% reinforcement condition, in which two dimensions were relevant and redundant in Stage 1, produced positive transfer.

References

Bourne, L. E., Jr. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, **54**, 201–207.

Bourne, L. E., Jr., & Haygood, R. C. The role of stimulus redundancy in concept identification. *J. exp. Psychol.*, 1959, **58**, 232–238.

Buss, A. H. Rigidity as a function of reversal and nonreversal shifts in the learning of successive discriminations. *J. exp. Psychol.*, 1953, **45**, 75–81.

Buss, A. H. Reversal and nonreversal shifts in concept formation with partial reinforcement eliminated. *J. exp. Psychol.*, 1956, **52**, 162–166.

Gagné, R. M., Baker, K. E., & Foster, H. On the relation between similarity and transfer of training in the learning of discriminative motor tasks. *Psychol. Rev.*, 1950, **57**, 67–79.

Gormezano, I., & Grant, D. A. Progressive ambiguity in the attainment of concepts on the Wisconsin Card Sorting Test. *J. exp. Psychol.*, 1958, **55**, 621–627.

Grant, D. A., & Curran, J. F. Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards. *J. exp. Psychol.*, 1952, **43**, 408–413.

Jenkins, W. O., & Stanley, J. C. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.

Kendler, H. H., & D'Amato, M. F. A comparison of reversal shifts and nonreversal shifts in human concept formation behavior. *J. exp. Psychol.*, 1955, **49**, 165–174.

Porter, L. W., & Duncan, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.

(Received November 9, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 376–383

VERBAL SATIATION AND CHANGES IN THE INTENSITY OF MEANING[1]

WALLACE E. LAMBERT AND LEON A. JAKOBOVITS

McGill University

Several investigators have demonstrated that *Ss* experience a change or loss of meaning for words which have been continuously repeated or fixated for a certain period of time (Basette & Warne, 1919; Mason, 1941; Smith & Raygor, 1956; Wertheimer & Gillis, 1956). For instance, Basette and Warne (1919) reported that the meanings of familiar nouns which were repeated aloud "dissipated" for their *Ss* within 3 or 4 sec. More recently, it was found that if *Ss* fixated a word exposed on a screen for 20 sec., their first association to the word is uncommon as measured by the Kent-Rosanoff Word Association Test (Smith & Raygor, 1956). The lapse or radical change of the meaning of a word as a result of its continued repetition is known as verbal satiation. Before the full implication of this concept for theories of learning and meaning can be determined, it is necessary to develop a method for reliably measuring the extent of meaning change which can be attributed to the satiation experience. The purpose of the present study is to extend the significance of the concept of verbal satiation by introducing a more comprehensive method of measuring the phenomenon and thereby relating it to a contemporary theory of meaning.

Osgood, Suci, and Tannenbaum (1957) have proposed an objective and reliable instrument for the measurement of certain aspects of connotative meaning. This instrument, the "semantic differential," consists of a series of scales each representing a 7-point, bipolar dimension. The meaning of a word, such as "father," is given by its position on an evaluative factor (its degree of goodness or badness), on an activity factor, and on a potency factor. The theory underlying this method assumes that the meaning of a word has a place in a multidimensional semantic space. A word without meaning would rest at the point of origin for all dimensions. Each scale has three degrees of polarity (see Jenkins, 1959) which describe the appropriate placement of the word along the scale. The middle position

TABLE 1

ILLUSTRATION OF A SEMANTIC PROFILE

Father							
Good	X	:	Y	:	:	:	Bad
Strong	:	:	:	Y	X	:	Weak
Passive	:	:	:	:	Y	X	Active
	3	2	1	0	1	2	3

Note.—*S* rates the concept "father" by indicating on each scale the position considered most appropriate. The polarity score in the present example is 6 for the X ratings and 2 for the Y ratings. The polarity-difference score for this example using one concept and three scales is −4, indicating that the second ratings (Ys) moved 4 scale units closer to the zero point.

[1] This research was supported in part by the Canadian Defence Research Board Grant D77-9401-10. We are particularly grateful to C. E. Osgood for his assistance and advice in the planning as well as the presentation of this study. R. C. Gardner and D. O. Hebb helped us by their suggestions and criticisms.

indicates lack of applicability of the word to either polar adjective, or "meaninglessness" on one dimension. A semantic profile is illustrated in Table 1.

In the present studies, the semantic differential is used as a method of indexing changes in meaning induced by means of verbal repetition of words. The study was carried out in several independent steps and the present report describes two separate experiments.

Experiment I

If *S* is requested to indicate the placement of a word on a semantic dimension immediately after continuous repetition of the word, his responses should fall on the zero point of the dimension if a total lapse of meaning has occurred. The semantic differential permits one to assess decreases of meaning short of total lapses. Thus, if under normal conditions *S* considers the word "father" and assigns it to Position 1 on an evaluative dimension ("extremely good") and then, after continuous repetition of the word (father) assigns it to Position 3 ("slightly good"), we will infer, when consideration is given to various controls, that the connotative meaning of the concept "father" has decreased (by about two-thirds) on this dimension.

Procedure

Three groups of college students were tested individually under several different conditions. As each *S* (with the exception of those in the retest control group, see below) came to the experimental room, he was asked to fill out a booklet in which he rated five words (child, me, rich, truth, family), each one on nine semantic scales (three scales for each of the three most prominent factors determined by Osgood et al., 1957). Then each *S*, depending on the condition to which he was assigned, followed a procedure outlined below. Each of the five words and each scale was printed on a separate 5 × 3-in. card. The cards were placed in a Kardex folder so that *E* could expose them in a predetermined order.

Experimental satiation.—For each of the 22 *Ss* in this group, a word was first exposed for about 1 sec. and *S* was asked to say the word aloud for 15 sec. at a rate of 2–3 repetitions per second. Then *E* immediately exposed a scale and *S* made his rating by pointing to one of the positions on the 7-point dimension. This was done for all words on all scales, a total of 45 responses per *S*. The order of presentation maximized the separation of the reoccurrence of each word and each scale. The initial and final ratings were subsequently compared.

Silence control.—The same procedure was used with 19 other *Ss* with the exception of one change: the *Ss* did not repeat aloud the words during the 15-sec. interval which elapsed between the time of exposure of the word and the semantic rating. They were initially instructed to "sit and wait" until a dimension was presented. No reference to "thinking about" the word presented was made in the instructions.

Different-word control.—The same 19 *Ss* participating in the previous phase also took part in this second control condition. Four additional words (war, death, teacher, athlete) were added to the booklet which was completed at the beginning of the experimental session. Using the same general procedure, the *Ss* repeated aloud a particular word during the time interval, but were then presented with a different word to rate. For example, the word "key" was exposed (the words used for repetition in this case were: key, moon, shoe, and book) and *S* repeated it aloud for 15 sec.; then the word "war" was exposed, read out loud by *S* who immediately gave a semantic rating on "war." This control condition was introduced to determine what effect the act of repeating words aloud had upon the stability of the ratings. Whereas the previous condition is a "unfilled-interval" control, the present condition is a "filled-interval" control.

Retest control.—The 22 *Ss* in this group did not fill out a booklet at the start of the experiment. In this case, ratings were taken immediately after the exposure of the words, with no repetition or interval interposed. Furthermore, after the first 45 ratings were obtained, the same complete series was repeated, and the ratings for the two series were compared.

Results

Mean polarity-difference scores were computed for each *S*. These represent the changes in degree of polarization (see Table 1) from the first testing under normal conditions and the second testing under experimental or control conditions. The changes are presented as average changes per word summed over all scales (in this case, nine). Thus, one *S* with a mean polarity-difference score of −3.6 had a total polarity score for the first testing of 102 (based on 45 ratings, 5 words on 9 scales) compared to 84 for the second testing under the satiation condition. The difference between these two totals, or 18, is 3.6 scale unit changes per word. A minus sign indicates a change from a higher to a lower score, i.e., a decrease in the intensity of association between the word and the bipolar adjective. Values of zero would indicate no change and positive scores would indicate an increase in intensity of connotative meaning.

Table 2 shows that the decrease in the intensity of meaning occuring under the experimental satiation condition was significantly greater than zero. Tests of significance between the experimental and control conditions on meaning change are presented in Table 3. The differences between the experimental and each of the three control conditions are significant, while none of the differences among the control conditions reaches significance. We conclude, therefore, that the decrease of meaning (verbal satiation) obtained with the experimental treatment is attributable to the continuous repetition of the word just before semantic ratings were made, and not to either of three other possible features of the experiment, namely, the unreliability of the measuring instrument, the 15-sec. interval period, or the interpolated task of repetition.

TABLE 2

AVERAGE CHANGE IN POLARITY PER WORD OVER THE SUM OF 9 SCALES: EXP. I

Condition	Change in Polarity			
	N	Mean	*SD*	*t*
Satiation	22	−2.85	2.93	4.45*
Silence control	19	0.03	1.41	0.09
Diff.-word control	19	−0.66	1.91	1.46
Retest control	22	−0.21	0.73	1.31

* Significantly different from zero beyond the .01 level.

TABLE 3

H TESTS OF SIGNIFICANCE BETWEEN CONDITIONS: EXP. I

Condition	Silence Control	Diff.-Word Control	Retest Control
Satiation	12.83	6.56	11.10
Silence control	—	0.92	0.69
Diff.-word control	—	—	0.03

Note.—If $H > 6.63$, $P \leq .01$; if $H < 3.84$, $P > .05$.

Examination of the data reveals that verbal repetition of a word may lead to small decreases of meaning as well as total lapses. In fact, when we calculate scores which each *S* would have obtained had there occurred a total loss of meaning, the obtained scores represent only 21% of those values that would have been obtained had there occurred a total loss of meaning as measured by the semantic differential. It is left for further studies to determine what amount of repetition would yield maximum satiation effects.

The results are consistent with Osgood's (1953, p. 410 ff.) interpretation of meaning. In this framework, the meaning of a symbol (or sign) is some replica of the actual reactions elicited by the environmental event which serves

as a referent for the sign. These mediating reactions should follow the same principles as any other response. During verbal repetition, the mediating reactions are repeatedly and rapidly elicited. Under such circumstances we would expect that a form of reactive inhibition would be generated which would temporarily decrease the availability of the mediators. Since Osgood's semantic differential is designed to index the strength as well as the kind of mediating reactions elicited by a sign, the development of inhibition during verbal repetition should be exhibited as a decrease in the extremeness of the ratings on the semantic differential, as indeed the present experiment has demonstrated.

The second experiment to be described was meant to replicate the above findings and to study further the nature of mediating reactions. One might argue that the changes in intensity of meaning noted above may be due in large part to the muscular components of the act of continuous verbal repetition. The following study bears directly on this point.

Experiment II

For Osgood, the mediation process is the meaning of a symbol. He leaves the question open as to the possible locus (loci) of mediational responses. They may be peripheral ("muscular or glandular reactions") (Osgood, 1953, p. 697) or central ("purely neural responses") (Osgood, 1957, p. 7). In this study, we presume that the mediation process that transmits significance to a symbol is inhibited by the continual verbal repetition of that symbol. The fact that the satiation effect was noted only for the experimental *Ss*, those who continuously repeated a symbol before it was rated, suggests that the mediation process may in large part be dependent on muscular reactions. In order to test the comparative importance of peripheral and central components of mediation processes, one might compare the behavior of the experimental *Ss* described above with another group of *Ss* who would be directed to "think about the word presented" but not to repeat it aloud. Should such a group display the satiation effect, however, one could still argue that they actually had said the words subvocally and no conclusive evidence would be given either the central or peripheral possibilities.

We attempted to circumvent this ambiguity by using another procedure which indirectly tests the comparative influence of peripheral responses and a central cognitive process on verbal satiation. We assume that saying aloud a meaningful word involves both muscular-glandular activity as well as some more central cognitive response as the meaningful nature of the symbol is registered. On the other hand, saying aloud a meaningless word with low association value involves peripheral muscular response accompanied by diversified cognitive activity, such as searching for possible significance in the word or for associations with the sound or form of the word, etc. In the case of the meaningful word, there is a relation between the peripheral and cognitive activities which is mediated by the meaning of the symbol. We argue that the cognitive activity is only distantly or not identifiably related to the peripheral in the case of the saying aloud of a meaningless word. The muscular reactions brought into play in the continual repetition of the words "canoe" and "nuka" (with the accent on the first syllable) are identical. (This assertion is supported by the fact that a listener cannot determine whether *S* is repeating "canoe" or "nuka," once the sequence of repetition is underway. The fact that the original peripheral feedback of the sequence "nu-ka-nu-ka" is different from that of the sequence "ca-noe-ca-noe" does not invalidate the present argument since the rest of the two sequences are essentially identical, and hence both should have the same effect on the peripheral responses.) The representational mediating processes which are elicited in the two situations, however, must be quite different since only one is a meaningful English word. The peripheral theory maintains that motor

responses are both necessary and sufficient for thinking. Thus, the continuous repetition of "nuka" and "canoe," involving identical muscular reactions, should have a similar satiation effect on the meaning of "canoe." For Osgood, whether the mediators are peripheral or central, they are *not* of the same form as the overt verbal response, since their character depends entirely upon the total reactions made to the thing signified and not the mediated verbalization. Since thinking of "ca-noe-ca-noe" and of "nu-ka-nu-ka" involve different cortical processes, repetition of "nuka" should not lead to satiation of "canoe."

Procedure

Twenty-three *Ss* (male and female public school teachers enrolled in a summer school) were tested under three different conditions. These are described below.

Satiation.—The procedure was identical with that used in the experimental satiation condition in Exp. I, except that words and scales have been changed. The purpose of this condition was to attempt to reproduce the results obtained previously as well as to serve as a comparison condition to the other two conditions described below.

Peripheral control.—The "centrality" hypothesis was tested by requiring *Ss* to repeat the words "grony" (accent on second syllable) and "nuka" before semantic ratings of the words "negro" and "canoe," respectively. The procedure was thus identical with the Different-Word Control in Exp. I. At the end of the experiment each *S* was asked whether he had "caught on" to the fact that "grony" and "nuka" were actually "negro" and "canoe" repeated backwards.

Nonsense control.—The effect of the repetition of a nonsense word on semantic ratings was determined by having *Ss* repeat "troga" and "blatu" before ratings of "house" and "soldier."

There were six words used in all (two words for each condition) and eight scales, six of these representing the three standard factors, and two scales representing a Familiarity factor (meaningful-meaningless; comprehensible-incomprehensible). The ratings given for the last factor were separately analyzed and will be discussed below.

Each *S* took part in three testings. The first was an assessment of *S*'s meaning of words under normal conditions. The words and scales were individually exposed in the Kardex folder and responses were recorded by *E* (no booklets were used in this experiment). The second testing consisted of the three conditions described above. The third testing, identical with the first, was given after a rest period of 5 min. to determine whether the satiation effect dissipates with time.

Results

The polarity-difference scores in Table 4 are averages for the group and represent mean changes of polarity per word on all six scales (absolute values cannot be compared with those in Table 2 since a different number of scales was used in Exp. I and II). Entries under the Exp. column are differences in polarity scores between the first and second testing, whereas entries under the Dissipation column are differences in polarity scores between first and third testings. It can be seen that the only significant changes in meaning took place under the Satiation condition. The possibility exists that the failure to obtain a satiation effect in the Peripheral Control and Nonsense Control conditions was due to the fact that the initial polarity of the particular words used under these conditions was already low, and hence could not decrease further. We therefore have calculated the mean

TABLE 4

AVERAGE CHANGE IN POLARITY PER WORD OVER THE SUM OF 6 SCALES: EXP. II ($N = 23$)

Condition	Experimental			Dissipation		
	Mean	*SD*	*t*	Mean	*SD*	*t*
Satiation	−1.95	2.06	4.43*	−1.76	2.11	3.91*
Peripheral control	0.19	1.37	0.65	−0.02	1.17	0.08
Nonsense control	−0.06	1.22	0.23	0.45	2.05	1.02

* Significantly different from zero at the .01 level.

TABLE 5

H TESTS OF SIGNIFICANCE BETWEEN CONDITIONS: EXP. II

Condition	Satiation, Exp.	Peripheral Control, Exp.	Nonsense Control, Dissip.
Satiation, Exp.	—	13.78	—
Satiation, Dissip.	0.21	—	10.14
Peripheral control, Dissip.	—	0.71	0.75
Nonsense control, Exp.	11.96	0.81	1.35

Note.—If $H > 6.63$, $P < .01$; if $H < 3.84$, $P > .05$.

initial polarity per word over the sum of six scales, as measured during the first testing, for all three conditions. These were as follows: 8.89 for the Satiation condition, 8.76 for the Peripheral Control and 9.17 for the Nonsense Control. None of these means was significantly different from another when a signed rank test was applied. It is evident, then, that differences in initial polarity of the ratings cannot account for the obtained results.

Table 5 shows that there is no significant difference between the Exp. and Dissipation scores under the Satiation condition, suggesting that the loss of meaning as a result of repetition persisted after a 5-min. rest (the mean dissipation score for the group obtained by subtracting the Exp. from the Dissipation scores is 0.19 ($t = 0.35$). The other differences in Table 5 are in harmony with a centrality hypothesis of meaning.

All but one of the 23 *S*s reported that they became aware of the fact that "grony" and "nuka" were actually "negro" and "canoe" repeated backwards. It appears from *S*s' reports that repetition of "grony," even though it involves the same muscular reactions, was not the same task as repetition of "negro." As one *S* put it, "When I was repeating 'grony' I was trying not to think of 'negro.' But when I was repeating 'father' I kept on thinking of 'father.' " Translating this into our terminology, we can say that, given instructions to repeat "grony," the mediating processes identified with "negro" were not consistently and reliably elicited. In view of these findings, we contend that in order to satiate the meaning of a symbol through continuous repetition one must consistently call into play some particular cognitive activity which is related to the symbol.

DISCUSSION

The continuous repetition of words reduces the intensity of their connotative meanings. The effect appears to be reliable and to persist for at least a period of 5 min. The effect also appears to depend on a consistent reactivation of some type of cognitive activity which is related to the word repeated. We conceptualize this phenomenon as a type of "semantic satiation" and consider the findings as support for a central interpretation of representational mediation processes.

The results show that by dint of continuous verbal repetition of words the semantic ratings *S*s made moved closer to the point of meaninglessness on the scales. It is not clear from the above results whether the satiation effect is restricted solely to the words which are repeated or if the effect is generalized, affecting the intensity of meaning of both the words and the bipolar adjectives, as well as the task of making a judgment. The finding of a decrease in intensity of meaning across the standard semantic scales is not inconsistent with the notion of a generalized satiation effect; if the bipolar adjectives were also satiated, they would have contributed to the decrease in the degree of association between the repeated words and the adjectives. If it could be shown, however, that *S*s were

also able to perceive increases in intensity of meaning on certain scales at the time of rating, we could argue that the inhibition effect is primarily restricted to the words which are repeated. As a test, we presented *S*s in Exp. II with two additional scales: "meaningful-meaningless" and "comprehensible-incomprehensible" and all words were rated along these scales as on the standard scales. These two scales should elicit a movement of ratings toward the "meaningless" and "incomprehensible" poles following the satiation treatment. Scores were assigned to *S*s for the amount of movement of ratings, comparing the first testing under normal conditions with the experimental and control condition testings, toward or away from the "meaningless" and "incomprehensible" poles. It was found that there was an average movement in the predicted direction of 1.43 scale units for the group under the satiation condition, a change which is reliably different from zero ($t = 3.25$, $P < .02$). No reliable change was noted for the other two control conditions ($t = 0.02$ for the satiation and peripheral control conditions and $t = 0.15$ for the satiation and nonsense control conditions).

One last point should be made about an alternative interpretation of our results; namely, that *S*s "caught on" to the fact that repetition of a word renders it less meaningful and consciously "played along" with *E* by making neutral judgments. There are several arguments against such an interpretation. Firstly, in the discussion which followed the experiment, *S*s were asked whether they thought that they had changed their judgments as the experiment went on and whether repetition influenced their ratings. The typical answers were "maybe" or "slightly," but the reason given for the change was ". . . because I forgot which judgment I had made previously." (In a current experiment using the same procedure, the question period at the end of the experiment was standardized and *S*s' answers were recorded. Only 1 out of 31 *S*s stated that repetition rendered the word more "meaningless." The typical answers were again "slight changes because I forgot what I had done before," or ". . . because I thought of something else.") Secondly, had *S*s wanted to please *E* by making neutral judgments, it is difficult to understand why they did not go all the way and give neutral ratings all the time. Thirdly, the dissipation scores in Exp. II exhibit the satiation effect on the words used under the Satiation condition only. Since the ratings of these words were mixed with those of the other conditions, it is improbable that *S*s could remember which were the satiation words and which were not. (In the current study mentioned above, only 3 *S*s could recall the words used for each condition.)

Summary

The phenomenon of verbal satiation—the decrease in the meaning of symbols—was studied by having college *S*s continuously repeat a word and then rate the word along scales of the semantic differential. Changes in semantic ratings, comparing normal and satiation conditions, indicate that there is a reliable movement of ratings toward the meaningless points of scales. Control group comparisons suggest that this movement is not due to the unreliability of the measuring instrument, to the time interval involved in repetition, nor to the activity of repetition per se. The phenomenon is conceptualized as a cognitive form of reactive inhibition and is related to Osgood's theory of representational mediation processes.

A second experiment sheds light on the nature of mediation processes, presenting evidence that they depend more on central than peripheral-muscular activities. In order to satiate the meaning of a symbol through continuous repetition, some particular cognitive activity which is related to the symbol must consistently be called into play. Further findings suggest that the satiation effect is not generalized but stems primarily from changes in the symbol which is continuously repeated.

References

Basette, M. F., & Warne, C. J. On the lapse of verbal meaning with repetition. *Amer. J. Psychol.*, 1919, **30**, 415–418.

JENKINS, J. J. Degree of polarization and scores on the principal factors for concepts in the semantic atlas study. Studies of Verbal Behavior, Rep. 1, Univer. Minnesota, 1959. (Mimeo.)

MASON, M. Changes in the galvanic skin response accompanying reports of changes in meaning during oral repetition. *J. gen. Psychol.*, 1941, **25**, 353–401.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

SMITH, D. E. P., & RAYGOR, A. L. Verbal satiation and personality. *J. abnorm. soc. Psychol.*, 1956, **52**, 323–326.

WERTHEIMER, M., & GILLIS, W. M. Some determinants of the rate of lapse of verbal meaning. Paper read at Rocky Mountain Psychol. Ass., Moran, Wyoming, 1956. (Mimeo.)

(Received November 11, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 384-390

DEPRIVATION AND REWARD MAGNITUDE EFFECTS ON SPEED THROUGHOUT THE GOAL GRADIENT

ROBERT FRANK WEISS

Columbus Psychiatric Institute, Ohio State University

Deprivation time and magnitude of reward are variables which have frequently been employed in studies of instrumental reward conditioning. A number of studies (e.g., Crespi, 1942; Logan, Beier, & Ellis, 1955; Zeaman, 1949) indicate that response speed is an increasing function of reward magnitude. Response speed is also an increasing function of deprivation time, up to a maximum point which appears to occur after about 2½ days of hunger (e.g., Barry, 1958; Besch & Reynolds, 1958; Cotton, 1953; Yamaguchi, 1951). Published studies of deprivation time have not been concerned with deprivation effects on the goal gradient, dealing only with starting speed, performance speed, or some combination of the two. There is a similar lack of information regarding reward magnitude effects on the goal gradient, with the exception of Crespi's (1942) study. In addition to this, no study of the combined effects of magnitude of reinforcement and deprivation time on response speed has yet been reported.[1] The purpose of the present study was to determine the combined effects of these variables on speed throughout the goal gradient.

In Hullian behavior theory (Hull, 1951, 1952; Spence, 1954, 1956) reward magnitude is one of the experimental operations defining incentive motivation (K), and deprivation time is one of the determinants of drive (D). The manner of combination of these intervening variables is a problem which has recently become a focus of theoretical analysis and research. While Hull (1951, 1952) originally suggested that D and and K combine multiplicatively to determine excitatory potential (E), Spence (1954, 1956) has recently hypothesized an additive relationship. Since response speed is considered to be a linear function of E, Hull's $D \times K$ hypothesis implies a multiplicative relation between the effects of deprivation and reward magnitude on response speed. Statistically this would appear as a significant interaction of deprivation and reward magnitude in which reward effects increased as deprivation increased. Alternatively, Spence's $D + K$ hypothesis implies the independence (nonsignificant interaction) of deprivation and reward effects. Presently available evidence favors the $D + K$ hypothesis. K has also been defined in terms of length of response chain (Spence, 1956) and delay of reinforcement (Spence, 1954), and both of these experimental variables have been shown to combine additively with deprivation time (Besch & Reynolds, 1958; Ramond, 1954).

The problem of deprivation and reward effects on the goal gradient is more complex. The goal gradient is nonmonotonic, speed first increasing and then decreasing, in a regular

[1] The writer has recently learned that William F. Reynolds and W. B. Pavlik have also completed a study of the joint effects of reward magnitude and deprivation, employing total running speed (starting plus running speed) as the dependent variable. Insofar as the two studies are comparable, they are in excellent agreement (personal communication from W. F. Reynolds).

manner, as the goal is approached (e.g., Hull, 1934; Weiss, 1958); and the effects of the interaction of deprivation or reward with distance from the goal are difficult to predict. This difficulty is exacerbated by a complex units-of-measurement problem (Logan, 1960) which is discussed later in the paper.

METHOD

Subjects.—The *Ss* were 72 naive male albino rats of the Sprague-Dawley strain, obtained from the Albino Farms in Redbank, New Jersey; they were between 100 and 140 days of age at the beginning of the experiment.

Apparatus.—The apparatus was the straight alley described in detail by Logan, Beier, and Ellis (1955).[3] The alley included an 8-in. startbox, a further 6 in. of starting area, and a 4-ft. runway, all of which were 4 in. wide and 4 in. high. The sidewalls were pine, lined with glass, and the floor was ¼-in. hardware cloth. Photoelectric cells, connected to Standard Electric timers, were mounted at the end of the starting section, and at 1-ft. intervals along the runway. There were six such timers, which measured (1) starting time—time from opening the start door to the first photobeam, (2–5) running time in the four successive 1-ft. sections of the runway, and (6) running time over the 4-ft. of runway, corresponding to the sum of Clocks 2 through 5. The timers were isolated in Celotex boxes mounted on sponge rubber inside second Celotex boxes. The entire alley was suspended in a large box, five sides of which were Celotex, the top being of glass covered with one-way vision cellophane. The alley was illuminated from within by five shielded 7-w. bulbs, mounted on one side of the alley, and by a 7-w. red bulb mounted over the foodcup. The room was darkened, windowless, and air-conditioned, with a hanging sound-tile ceiling.

Design.—The levels of food deprivation employed were 2, 12, 24, and 46 hr. The reinforcement magnitude conditions were 1, 3, 6, and 12 pellets. The uniform food pellets weighed .045 gm. each, and were prepared by the P. J. Noyes Co., Lancaster, New Hampshire, primarily of Purina. The design is essentially an extension of the Perin-Williams design (Perin, 1942). The Perin-Williams design represents a selection of one complete column and one complete row from a larger factorial design. The design employed here consisted of two complete columns and two complete rows (12 cells) selected from a 4 × 4 factorial design, as shown in Table 3.[5]

Procedure.—All *Ss* were placed on the 48-hr. feeding rhythm six weeks before the experiment began, during which time they were handled and weighed every second day. The *Ss* were allowed to eat as much as they could for 40 min., every second day. A synthetic diet was employed in order to prevent loss of *Ss* due to the stringent deprivation cycle. The diet was constituted as follows: 50% ground Purina Lab Chow, 30% lard, 20% B-3-F casein, plus vitamin and mineral supplements added shortly before feeding.[6] The rats gained weight slightly throughout the experiment.

The *Ss* were run one trial every second day. The reward was delivered with an audible click when the last photobeam was broken, assuring immediate secondary reinforcement even on the occasional trials when *S* did not immediately consume the reward. Each *S* was placed in an 18 × 18 × 4-in. gray delay box, immediately before and after running, for the length of time it took the preceding or following *S* to complete its trial. An equal diet was maintained, despite differences in the amount of food administered as reward, by means of a compensatory feeding of reward pellets approximately 1 hr. after running. The *Ss* were run for 55 trials, at which time speed, measured over the entire alley, was asymptotic, as indicated by a *t* test

[3] The *Ss* were run while the writer was at Yale University; this phase of the research was supported in part by a grant to F. A. Logan by the National Science Foundation. The writer wishes to thank Logan for his kindness and for helpful advice and criticism.

[5] This design, like that of Perin, was intended for curve-fitting (or, more accurately, surface-fitting), though the results have not been sufficiently regular to warrant the use of so elaborate a technique. The design is not as conveniently suited to analysis of variance as it is to curve-fitting, and it has been necessary to analyze separately the data from the two complete rows and the two complete columns.

[6] Abdec multiple vitamin drops and the following water solutions: $CuSO_4$ (.125%), $MnSO_4$ (5%), Fe_2SO_4 (5%); each administered in the amount of 1 drop/16.5 gm. diet. The writer wishes to thank J. Barboriac and W. Krehl for suggesting the constituents of the diet.

of the mean of the differences between the last 10 trials and the preceding 10 trials ($t = .37$, $df = 71$). They were then reassigned to other conditions, so as to constitute a second "replication" of the experiment, in a random order with the following restriction on randomness: (*a*) no *S* served twice in the same cell or drive condition, and as few as possible served in the same reward condition; (*b*) not more than one *S* from a given first-replication cell was assigned to any one second-replication cell. The *S*s were run for 16 trials under the new conditions, during the last five of which the new asymptotes were well established.

Results and Discussion

Starting speed.—Mean starting speed for each *S* was obtained over a block of five trials. Trials 51–55 of the first replication and 12–16 of the second replication were used. The data from the two complete columns and the two complete rows were analyzed as separate $4 \times 2 \times 2$ factorials (Table 1).[5] In both analyses, the replication effect was nonsignificant. The Replication × Reward and Replication × Deprivation interactions were found to be nonsignificant in both analyses; hence, data from the two replications were combined for further analysis. Mean starting speed for the combined replications is shown in Table 2; $N = 12$ for each cell.

TABLE 1

Starting Speed: Analyses of Variance

Source	Two Complete Columns			Two Complete Rows		
	df	*MS*	*F*	*df*	*MS*	*F*
Deprivation (*D*)	3	.520	4.33**	1	.539	4.49*
Reward (*K*)	1	2.460	20.50***	3	.351	2.93*
Replications (R)	1	.194	1.62	1	.280	2.33
$D \times K$	3	.106	.88	3	.153	1.28
$D \times R$	3	.117	.97	1	.137	1.14
$K \times R$	1	.381	3.17	3	.019	.16
$D \times K \times R$	3	.271	2.26	3	.038	.31
Error (within)	80	.120		80	.120	
Total	95	.165		95	.129	

* $P < .05$.
** $P < .01$.
*** $P < .001$.

TABLE 2

Starting Speed and Running Speed

Deprivation (Hours)	Magnitude of Reward (Pellets)			
	1	3	6	12
	Mean Starting Speed at Asymptote[a]			
2	.256			1.217
12	.570	.548	.855	.810
24	.512			.801
46	.871	.943	.768	1.147
	Mean Running Speed (Ft/Sec) for Trials 1–20[b]			
2	.78			.94
12	1.04	.92	1.22	1.14
24	.87			1.24
46	1.16	1.19	1.36	1.78

[a] Replications 1 and 2 combined, $N = 144$.
[b] Replication 1 only, $N = 72$.

As shown in Table 2, starting speed was an increasing function of reward magnitude at each of the four levels of deprivation. This effect was found to be significant in the analysis of the two complete columns ($P < .001$) and the two complete rows ($P < .05$). The deprivation effect was also found to be significant in the analyses of the two complete columns ($P < .01$) and the two complete rows ($P < .05$).[6] Spence's (1954, 1956) assumption of an additive relation between *D* and *K* implies a nonsignificant interaction

[5] A regular factorial design was somewhat conservative in this case, since it did not take advantage of the fact that the "replications" variable involved repeated measurements on the same *S*s. Since each *S* served in only 2 of the 12 conditions, however, it was not possible to take account of this correlation by means of a Treatments × *S*s type of design.

[6] As shown in Table 2, the 12-hr. groups ran somewhat faster than the corresponding 24-hr. groups. This difference was not significant.

between deprivation and reward magnitude, differences among reward levels being independent of level of deprivation. Hull's (1951, 1952) multiplicative assumption implies a significant interaction in which reward effects increase as deprivation increases. The starting speed data offer no support for a multiplicative hypothesis. Reward effects do not appear to depend systematically on deprivation among the 46-, 24-, and 12-hr. groups. When the 2-hr. group is also considered, there is a suggestion of an interaction opposite to that predicted by Hull; reward effects appear greater at the lowest level of deprivation than at the higher levels. The Deprivation × Reward interaction (including the apparent covergence of the 2-hr. group with the others) was found to be clearly nonsignificant in the analyses of the two complete columns ($F = .88$; $df = 3/80$) and the two complete rows ($F = 1.28$; $df = 3/80$). Deprivation time and reward magnitude thus appear to combine additively to determine starting speed, as is implied by Spence's assumption of an additive relationship between drive (D) and incentive motivation (K).

Running speed.—Reward effects on asymptotic running speed fell short of significance,[7] rendering inconclusive the nonsignificant Deprivation × Reward interaction. In an effort to clarify the running speed results, an analysis was conducted on the data from Trials 1–20. Mean running speed (in ft/sec) on Trials 1–20 is shown in Table 2. Despite the fact that these data were necessarily restricted to the first replication alone, the effect of reward magnitude was found to be significant in the analyses of the complete columns ($F = 6.87$; $df = 1/40$; $P < .025$) and the two complete rows ($F = 3.10$; $df = 3/40$; $P < .05$). The deprivation effect was also found to be significant in the analyses of the two complete columns ($P < .01$) and the two complete rows ($P < .025$). The data on running speed in Table 2 appear to indicate a tendency for reward effects to increase with level of deprivation. The Deprivation × Reward interaction, however, was not significant in either the analysis of the two complete columns ($F = 1.02$; $df = 3/40$) or the two complete rows ($F = 1.42$). These results, while not conclusive, do tend to suggest that deprivation time and reward magnitude combine additively to determine running speed.[8]

Goal gradient.—Figure 1 depicts the development of the goal gradient during the course of acquisition. The N for each curve is 72. The speed gradients exhibit the nonmonotonic shape found by Hull (1934). In contrast to Hull's data, however, it can be seen that the gradient becomes more, rather than less, peaked as the number of trials increases. This increase in peakedness of the speed gradient during acquisition is in agreement with data recently

[7] It should be noted, however, that this result is not in agreement with the results of the two other studies conducted in the same apparatus, using the same standardized brand of reward pellets and employing running speed as the dependent variable. Logan, Beier, and Ellis (1955) found that their 1- and 9-pellet reward conditions differed significantly at asymptote (62 trials), as did Beier's (1958) 1-, 7-, and 13-pellet groups (Trials 78–93).

[8] It may be recalled that while there was an apparent tendency for reward effects on running speed to increase with increasing deprivation, the opposite tendency appeared to be present in the starting speed data. If these effects had been significant, starting speed would be determined by the exponential summation (+) of deprivation and reward, and running speed by their multiplication.

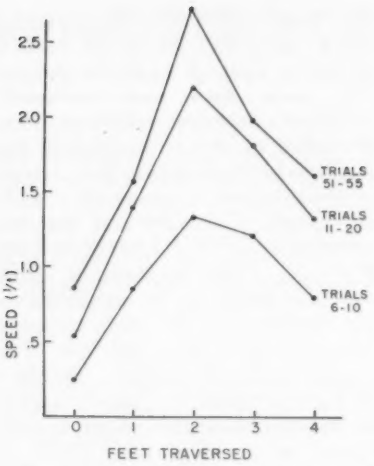

FIG. 1. Goal gradient at three levels of training: Trials 6–10, 11–20, and 51–55. (Zero feet traversed indicates starting speed.)

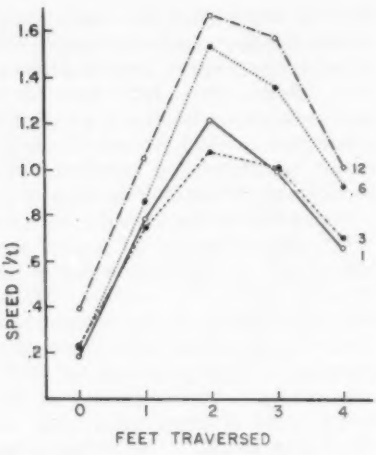

FIG. 2. Goal gradient at four levels of reinforcement magnitude.

obtained by the writer, using a bar-pressing apparatus (Weiss, 1958). Figure 1 also indicates that there is a change in the acceleration of both the ascending and descending slopes of the goal gradient during the course of acquisition. On Trials 6–10 the ascending slope of the goal gradient appears concave downward, this trend appearing more clearly in the descending slope. The intermediate goal gradient is more nearly linear, while each of the two slopes of the asymptotic gradient is concave upward.

The goal gradient at four levels of reward magnitude is shown in Fig. 2. Trials 6–10 (first replication only) were employed. (As is clear from the preceding sections, the goal gradient was most marked at asymptote, but deprivation and reward effects on speed of running were greater earlier in acquisition, and earlier acquisition trials have therefore been used here.) The effect of reward magnitude on speed appears to be greater in the middle of the alley than the start or end. These results are not in agreement with those of Crespi (1942), which appear to indicate a decreasing steepness of the goal gradient as reward magnitude increases. The goal gradient at four levels of deprivation is shown in Fig. 3. The effect of deprivation on speed appears to be more pronounced in the

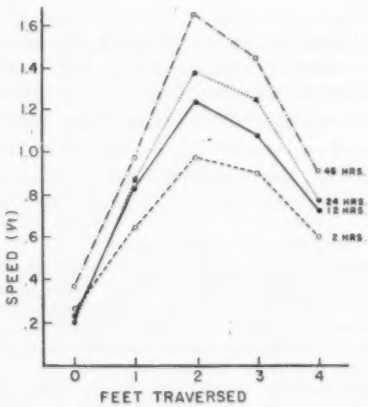

FIG. 3. Goal gradient at four levels of deprivation.

middle of the alley than at the start or end.

Since Hull's (1934) initial study of the goal gradient, two factors have been regarded as contributing to gradient effects: (*a*) within-chain delay of reinforcement, and (*b*) generalization of stopping from the termination of the response chain to earlier points in the chain. A recent study (Weiss, 1958) appears to indicate that the inhibitory (stopping) variable is operative even when (as in a bar pressing situation) there is no physical necessity for *S* to slow down gradually before stopping completely. In addition to this aspect of inhibition, the inhibition generated by the work involved in responding may be regarded as accumulating from the onset of the response until its termination (Hull, 1943). Two variables, operating in opposite directions, thus appear to be involved in the determination of the goal gradient: incentive (*K*) and inhibition (*I*). This interpretation is supported by the results of studies in which distance from the goal was varied between groups by varying the length of the behavior chain to be learned and measuring speed at the beginning of each chain. Weiss (1958) varied both work and response length and found an effect of duration of effort (inhibition) which appeared to be independent of any effect of within-chain delay of reinforcement. Studies employing straight alleys of markedly different length obtained effects which appear to be due to delay. (Besch & Reynolds, 1958; Hull, 1934) though Fredenburg (1956) did not obtain speed differences in a study involving a smaller difference (4 vs. 11 ft.) in alley length. It may be noted that goal gradient effects are obtained even when the work involved in the response is too low to produce between-groups work-duration effects (Weiss, 1958) or when, as in the present study, the distance between the various points in the alley is not as great (absolutely or relatively) as would be required to produce between groups delay effects.

If the simplifying assumption were made that the ascending portion of the goal gradient represents the effects of delay, relatively uncontaminated by inhibition, the following interpretation of the data reported in this study could be made. The diverging gradients at different levels of training might reflect an interaction between delay of reinforcement and number of trials, which would be in accord with the theoretical assumption of a multiplicative relation between incentive motivation (*K*) and habit strength (*H*). The diverging gradients at different levels of deprivation would then imply a multiplicative relation between within-chain delay of reinforcement and deprivation time, which is not in agreement with the theoretical assumption of an additive relationship between *D* and *K*. Of greater importance, however, is the contradiction between this interpretation and the results of Besch and Reynolds (1958) who found no interaction between deprivation and alley length. The simplifying assumption made above would thus appear to be untenable.

The interpretation above also involved a second assumption—that speed bears an *identical* relationship to *E* in each section of the alley. This assumption is called into question by a recent exploratory study (Logan, 1960), employing Hull, Felsinger, Gladstone, and Yamaguchi's (1947) adaptation of paired-comparisons scaling procedure for quantifying *E*. The relationship between speed and *E* appears to be linear in each section of the alley, but the constants in the linear equations differ. If, for example, $E = 2 \times$ Speed in one part of the alley, and $10 \times$ Speed in another part, the gradient of *E* could very easily differ in shape from the gradient of speed. The present indications (Logan, 1960) are that the gradient of scaled *E* is a monotonic decreasing function of distance from the goal. Adequate theoretical interpretation of goal gradients of speed must, therefore, wait on a satisfactory solution of the problem of appropriate units-of-measurement, as well as a separation of incentive and inhibition effects.

Summary

The effects of deprivation time and reward magnitude on speed in the five segments of a straight alley were studied by means of a modified 4 × 4 factorial design. Deprivation time and reward magnitude appeared to combine additively (significant main effects, nonsignificant interaction) to determine starting speed, as is implied by an assumption of an additive relationship between drive (D) and incentive motivation (K). Reward effects on asymptotic running speed fell short of significance, though an analysis of earlier trials did tend to suggest that deprivation and reward magnitude also combine additively to determine running speed. Reward effects on running speed (Trials 1–20) appeared to increase as deprivation increased, but this interaction did not approach statistical significance. The goal gradient was found to increase in steepness during the course of acquisition, and the effects of deprivation and reward were both greater in the middle portions of the alley than at the start or end. The gradient data were discussed in terms of the effects of within-chain delay of reinforcement, duration of effort, and generalization of inhibition, as well as in relation to a problem of appropriate units of measurement.

References

Barry, H., III. Effects of strength of drive on learning and on extinction. *J. exp. Psychol.*, 1958, **55**, 473–481.

Beier, E. M. Effects of trial-to-trial variation in magnitude of reward upon an instrumental running response. Unpublished doctoral dissertation, Yale Univer., 1958.

Besch, N. F., & Reynolds, W. F. Alley length and time of food deprivation in instrumental reward learning. *J. exp. Psychol.*, 1958, **56**, 448–452.

Cotton, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.

Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

Fredenberg, N. C. Response strength as a function of alley length and time of deprivation. Unpublished doctoral dissertation, State Univer. Iowa, 1956.

Hull, C. L. The rat's speed of locomotion gradient in the approach to food. *J. comp. Psychol.*, 1934, **17**, 393–433.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Hull, C. L., Felsinger, J. M., Gladstone, A. I., & Yamaguchi, H. G. A proposed quantification of habit strength. *Psychol. Rev.*, 1947, **54**, 237–254.

Logan, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

Logan, F. A., Beier, E., & Ellis, R. A. The effect of varied reinforcement on speed of locomotion. *J. exp. Psychol.*, 1955, **49**, 260–266.

Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

Ramond, C. K. Performance in instrumental learning as a joint function of delay of reinforcement and time of deprivation. *J. exp. Psychol.*, 1954, **47**, 248–255.

Spence, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In, *Kentucky symposium: Learning theory, personality theory, and clinical research.* New York: Wiley, 1954. Pp. 1–21.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Weiss, R. F. Response speed, amplitude, and resistance to extinction as joint functions of work and reinforcement ratio. Unpublished doctoral dissertation, Yale Univer., 1958.

Yamaguchi, H. G. Drive (D) as a function of hours of hunger (h). *J. exp. Psychol.*, 1951, **42**, 108–117.

Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–484.

(Received November 14, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 391–403

SOME DETERMINANTS OF TWO-DIMENSIONAL VISUAL TRACKING BEHAVIOR[1]

JACK A. ADAMS AND LOUIS V. XHIGNESSE

University of Illinois

In the beginning it is common for psychological research on a topic to receive its impetus from problems and observations of everyday human activities. As the research progresses, a process of simplification and abstraction ensues where investigators seek isolated and controlled manifestations of the phenomena to observe uncontaminated actions of one or more variables. In pursuing this traditional course of law-finding, scientists become increasingly removed from the complex situations which first initiated the observations. The general presumption is that our scientific understanding of complex behavior will tend to emerge as more and more of the laws derived in simple situations are uncovered. While this presumption is essentially correct, it must be kept in mind that a complex phenomenon is more than an aggregate of simple phenomena because it can involve new variables and interactions among known variables that are absent in simpler contexts. Laws governing simple phenomena cannot necessarily be assembled in an a priori fashion to unequivocally explain complex activity. Instead, they must be examined empirically within the complex situation itself to determine whether new factors and relationships are present. While a few investigators always show a steady interest in research at the level of the complex domain, it seems safe to say that research on relatively restricted, simple domains dominates the total effort for a long time. Only after a solid substructure of research has been built does a more intensive attack begin on behavior in complex situations.

Research on tracking behavior is following this traditional progression. Mainly, the goad for laboratory research in tracking arose out of complex applied tasks where a human operator was a tracking element in a closed-loop system and had the job of continuously nulling a displayed error quantity. Several hundred experiments, mostly since World War II, document the concern with tracking (Lockard, 1956), and much of it is basic in orientation. Almost all of this research has been done on simple one-dimensional tasks where *S* attends to a single stimulus source with unidimensional variation of events and responds with a control system having a single dimension of movement. Recently, however, there has been more attention given to variables in the organization of complex tracking tasks having two or more stimulus sources and a corresponding number of dimensions of the control system for response to them (Charipper, 1959; Fitts & Simon, 1952; Jackson, 1958; Sweeney, Bailey, & Dowd, 1957). Uncovering relationships for these more complex systems should proceed at a good rate because the research will be grounded in a matrix of knowledge about the one-dimensional component tasks that comprise the whole. Some of our present

[1] This research was supported by the United States Air Force under Contract No. AF 49(638)-371, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

research needs lie in further developing an understanding of the interaction of simple effects and the nature of new variables in these more complex tasks (Adams, in press). The experiment to be reported here is broadly directed toward this problem in two-dimensional visual tracking.

One of the primary new variables in two-dimensional visual tracking, absent in one-dimensional visual tasks, is the spatial separation of stimulus sources which places a requirement for an observing or scanning response (Adams, in press; Senders, 1952; Wyckoff, 1952) to scan the two stimulus sources and detect error. When stimulus sources are close together the demands on the observing response are low because *S* needs only to orient his eyes to each source or, conceivably, might be able to apprehend two different events simultaneously in his visual field and differentially respond to them without scanning eye movements. But, as the sources become widely separated, *S* must use head and eye orientations to successively sense events at each source. It would seem reasonable for response proficiency to decrease as the degree of spatial separation is increased because the wider the separation the greater the likelihood that error will go uncorrected while *S* is momentarily watching the other source. However, there are at least two other variables which might be expected to interact with spatial separation. One of these is the degree of coherency or predictability of the stimulus events in each source. A human *S* has a well-known and potent capability for learning the dependencies among events and to develop *perceptual anticipation* (Bartlett, 1951; Poulton, 1952, 1957). In the absence of anticipation, *S* will wait for stimulus change to occur before initiating response which means that he will have error as a function of at least one reaction time interval. But, with anticipation, the response movement will be initiated momentarily before the change and the control system will be in the correct position at the time the change actually occurs. To be identified as beneficial anticipation the interval between stimulus and response onset must be less than the expected reaction time interval (Helson, 1949; Poulton, 1952). Ideal anticipation would be perfect time coincidence of stimulus and response. If both events appearing simultaneously in two sources are predictable *S* can respond to them with minimum or even no visual sensing at the moment of response because he "knows" what the correct responses should be. If only one of the sources has a predictable event *S* still has an advantage because the observing response need orient the sensors only to the uncertain event, thus eliminating the rather pressing demand of successively sensing both events as rapidly as possible. However, if coherency is low in both sources and there are few or no regularities to be learned, *S* has no choice but to maintain a highly active scanning of both sources in an attempt to sense event changes as soon as they occur.

The other variable which might be expected to interact with spatial separation is the speed of event change at each stimulus source. When the events change rapidly *S* stands a good chance of missing some of them or sensing them late because his observing response cannot keep up with all of the stimulus activity. Anticipatory behavior might emerge here as an important compensatory

mechanism to somewhat offset the particularly high demands that the task is imposing on *S*.

The experiment to be reported was designed to explore these matters with a $2 \times 2 \times 2$ factorial design where two values of spatial separation, stimulus coherency, and speed of event change were studied in a two-dimensional discrete visual tracking task.

Method

Apparatus.—The experimental task was the two-dimensional discrete tracking apparatus, a pursuit tracking task with two visual stimulus sources. Each stimulus source had three discrete stimulus lights which were automatically programmed and *S* had a three-position control stick in each hand for response to them. Beneath each stimulus light was a feedback light which informed *S* whether he had the control in the correct position. As in any pursuit tracking task, *S* responded to keep the stimulus and feedback indicators aligned as much as possible and keep the error discrepancy between them to a minimum. The control stick for each hand was 2 in. high and was mounted on a broad, flat arm rest of the chair in which *S* was seated. The control stick was attached to a three-position switch which returned the control to the center position when released. Spring leaf contacts on the switch minimized the feeling of discreteness that is ordinarily present in a switch. The control was essentially free and easy to operate, although there was a slight spring leaf resistance as the control was moved to either position from center. Movements of the control and the displayed feedback lights were completely compatible. The left-hand control was for the stimulus source on the left and the right-hand control was for the source on the right. The line of control movement was parallel with the movement of the feedback lights. Furthermore, the directional movement of the feedback lights was the same as the directional movement of the control. The general intent for the control system was to make it as simple as possible.

The two stimulus sources were at a distance of 32 in. from *S* and at eye level on a 180° semicircular metal track. The stimulus sources could be moved along the track to attain any amount of spatial separation up to 180°. Each stimulus source was a small metal box with a front surface of $2 \times 3\frac{1}{4}$ in. The source consisted of two horizontally arrayed rows of lights. The three stimulus lights were in the top row and were red, white, and green jeweled lamps of $\frac{1}{4}$ in. diameter, with the white light in the center. Directly under each stimulus light was a feedback light which was a 1/25-w. neon indicating lamp Type NE-2 and glowed distinctive yellow-orange when illuminated. For both sources, the presentation of stimulus events and the timing of their duration was programmed on five-hole punch tape and read by a motorized tape reader, Model 2, manufactured by the Commercial Controls Corporation. Everything, including inter-trial rest periods, was programmed on the tape so that an entire experimental session ran off without intervention from *E*.

Description of stimulus series.—Because of hypotheses about the role of perceptual anticipation, a coherency variable was specified by two sets of probability tables which defined a high and a low degree of statistical organization for the stimulus events. Table 1 gives the zero-order and conditional probability tables for each degree of stimulus organization. The redundancy was 68% for the high-coherency series and 37% for the low-coherency series (Attneave, 1959). For any particular experimental

TABLE 1

Zero-Order and Conditional Probability Tables for the Two Coherency Levels of Stimulus Organization

High Coherency

		$p(i)$		$pi(j)$	j Red	j White	j Green
	Red	.25		Red	0	1.0	0
i	White	.50	i	White	.50	0	.50
	Green	.25		Green	0	1.0	0

Low Coherency

		$p(i)$		$pi(j)$	j Red	j White	j Green
	Red	.33		Red	.50	.50	0
i	White	.33	i	White	.50	0	.50
	Green	.33		Green	0	.50	.50

Note.—The conditional tables give the probability of a jth event following an ith event.

condition, the events in each of the two sources used the same set of probability tables, but the events for each source were independently generated. The rate of events was either 1 per sec. or 1 per 2 sec., which is the amount of time that an event remained on before change and is similar to cycles per second for, say, a sine wave input in continuous tracking because it determines the number of response changes per unit of time. There was no dead time between events. They were perfectly contiguous in time.

Table 1 shows that for a source in the high-coherency case no event follows itself and every colored event is always followed by a white event. However, a white event has a .50 probability of being followed by red or by green. Such a series for a source might look like red, white, red, white, green, white, red, white, etc. In the low-coherency case, however, a colored event has a probability of .50 being followed by itself or by a white event. This meant that a colored event, in sometimes repeating itself, persisted for longer than the 1 or 2 sec. defined as the duration of each event for any particular group. Thus, on the occurrence of three successive red events for a 2-sec. condition, *S* would be confronted with a requirement in a source to hold the control aligned with red for 6 sec. A low-coherency series in a source might look like white, red, red, white, green, green, white, red, red, red, etc. All events are treated as discrete and of fixed duration for the purposes of constructing the probabilistic, time-based series but, by allowing an event to repeat itself in the low-coherency case, *S* sees some events as being longer than others. Thus, in the low-coherency case the within-source instability also gives low between-source coherency because *S* is never certain when a colored event will change in either source. This is to be contrasted with the high-coherency series where the within-source regularities give high between-source coherency because a colored event in the left source is always paired with a white event in the right-hand source, and both of them simultaneously change to a white event in the left source and a colored event in the right source. In the high-coherency series *S* always has an event class present in one of the stimulus sources which he can potentially predict in time and type and reduce the requirement for visual scanning, while in the low-coherency series he never has this stability. Perhaps a better way of characterizing it is in terms of the type of certainty which is present for the various classes of events. In the high-coherency series, the white events have *event-time certainty* because white invariably follows a colored event after a fixed presentation time for the colored event. Both the type of event and the time of occurrence are rigidly established. The colored events in both the high and the low series have *time certainty* because the white events preceding them always have a fixed duration, although it cannot be specified whether the color of the next event will be red or green. A white event in the low series has *event certainty* because it always follows a colored event, but the time of occurrence is uncertain because the preceding colored event has varying duration. The nature of the task provides some confounding of time uncertainty and event uncertainty in the two levels of coherency. Within a level of coherency, we are unable to independently manipulate event and time certainty.

A remark is in order on the design of the probability tables. The restriction of a white event always following a colored event gives the same constraints to the stimulus events that are inherent in the mechanical design of the control system. It is impossible for the control stick to pass from a red event to a green event without first passing through white, and so it is necessary to impose the same transitional constraints on the stimuli. Otherwise the stimulus states could have sequences (e.g., red, green, red) which could not be followed by the control system without error. By having the same constraints apply to both stimuli and control, the stimuli do not take action which, in principle, cannot be matched by the control system. Failure to match can then be interpreted in terms of behavior instead of the task design.

Response measurement.—The overall measure of proficiency used was simultaneous time on target, which is the amount of time on a trial that *S* had the stimulus lights and feedback lights simultaneously aligned in both stimulus sources. In addition, an Esterline-Angus operations recorder was used on selected trials to record all stimulus and response events. A detailed analysis was made of the chart records taken on the final trial (50).

Experimental design.—A $2 \times 2 \times 2$ factorial design was used, with 12 *Ss* randomly assigned to each of the eight groups. The two levels of stimulus coherency were *high* and *low*, as previously described. Spatial separations were *adjacent* and *separated*. The adjacent condition had the stimulus sources symmetrically aligned 5° on either side of the median plane of *S*'s visual field. For the separated condition the symmetrical alignment of sources was 30° from the median

plane. In the adjacent condition the two small rectangular boxes containing the lights and serving as the stimulus sources were virtually touching, whereas in the separated condition they had a disparity of 60° on the semicircular track. The two values of the rate variable were *1 sec.* and *2 sec.* The designation of each group and its experimental treatment is as follows:

Group	Treatment
HA-1	High coherency, adjacent, 1-sec. rate
HA-2	High coherency, adjacent, 2-sec. rate
HS-1	High coherency, separated, 1-sec. rate
HS-2	High coherency, separated, 2-sec. rate
LA-1	Low coherency, adjacent, 1-sec. rate
LA-2	Low coherency, adjacent, 2-sec. rate
LS-1	Low coherency, separated, 1-sec. rate
LS-2	Low coherency, separated, 2-sec. rate

Each *S* had 50 practice trials distributed over three experimental sessions on three different days. Twenty trials were administered on each of the first two days, and 10 trials on the third day. A trial was 100 sec. in duration. Intertrial rest was 40 sec. Two program tapes (high- and low-coherency) for 1-sec. groups presented 2,000 events at each stimulus source during a 20-trial experimental session, with the 2,000 events in each source being independently generated in accordance with the probability structure shown in Table 1. The two program tapes for the 2-sec. groups presented 1,000 events at each source during a session, and these were the first 1,000 events of the 2,000 given to the 1-sec. groups. The same program tapes were used for each session. Instructions emphasized the totality of the task and the importance of keeping both stimulus sources in simultaneous alignment as much as possible.

Subjects.—The 96 *S*s were undergraduate university men. They were paid for their participation.

Results

Figure 1 presents the performance curves for the eight groups. Mean simultaneous time on target is plotted against trials. The plotted point value is the mean of two trials. The breaks in the curves represent the intersession intervals. A Lindquist Type I trend analysis design (Lindquist, 1953) was performed on the simultaneous time on target measure

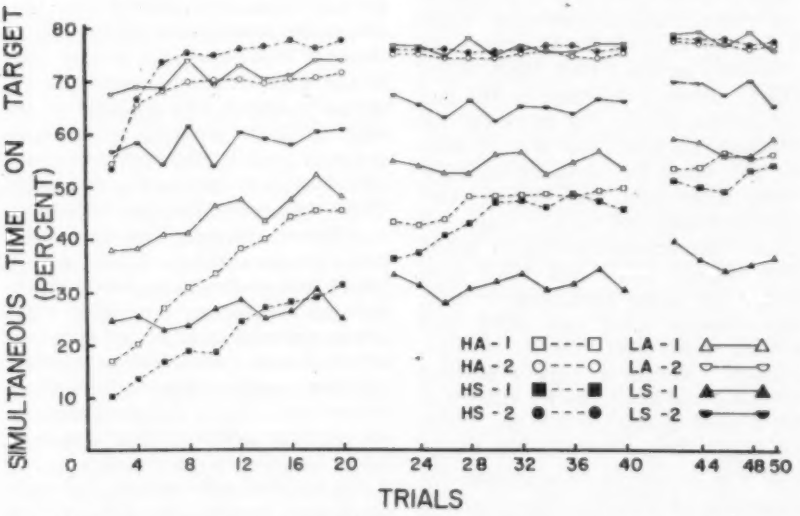

FIG. 1. Mean percentage time that the two stimulus sources were being simultaneously aligned. (For ease of interpretation, all high-coherency groups are coded with a broken line between symbols, all low-coherency groups with a solid line between symbols, all adjacent groups with open symbols, all separated groups with solid symbols, all 1-sec. groups with straight-line symbols, and all 2-sec. groups with curved line symbols.)

for all eight groups over the final 10 trials[2] and revealed, what is obvious from Fig. 1, that the experimental treatments were highly potent determiners of final proficiency level. A null hypothesis test of Methods variance was significant beyond the .01 level, as was the Trials by Methods interaction variance in a test of the null hypothesis for the trends of trial means.

Because of the number of treatments, it will clarify the exposition to examine the effects of each of the variables separately. The analysis is concentrated on terminal levels of proficiency.

Effects on Simultaneous Time on Target

High coherency.—One of the most significant findings is that regularities of the high-coherency stimulus series result in learned behavior which largely eliminates effects of the spatial separation of stimulus sources. While a Type I analysis for the four high-coherency groups (HA-1, HS-1, HA-2, HS-2) allows a rejection of the null hypothesis at the .01 level for the Methods variance and the .05 level for the Trials by Methods variance, these differences are almost completely determined by the rate variable. Examination of Fig. 1 shows that Groups HA-2 and HS-2 have about the same performance level, and Groups HA-1 and HS-1 also exhibit equivalent levels of proficiency. Neither a Type I analysis for HA-2 and HS-2, nor for HA-1 and HS-1, gave significant F ratios at the .05 level. It can be concluded that with a given rate of event change the separation of stimulus sources makes no difference for performance as long as the stimulus series has high coherency.

Low coherency.—In contrast to the high-coherency groups, the low-coherency groups are fully differentiated on all other variables. Figure 1 shows that among the four low-coherency groups (LA-1, LS-1, LA-2, LS-2) the 2-sec. rate results in higher performance than the 1-sec. rate. Within any rate condition, the separated stimulus sources produce poorer performance than the adjacent ones (Fitts & Simon, 1952). In the absence of stable regularities in the stimulus series, performance is quite sensitive to the effects of the other variables in the experiment. A Type I analysis for the four low-coherency groups allowed rejection of the null hypothesis for the Methods variance and the Trials by Methods variance at the .01 level.

Figure 1 also shows that the forms of the acquisition curves for low-coherency groups are different from those of high-coherency groups. All of the low-coherency curves are essentially linear with gradual slope, while the high-coherency curves have the negatively accelerated curvilinear characteristics commonly associated with visual-motor learning curves.

Adjacent sources.—The four adjacent groups (HA-1, LA-1, HA-2, LA-2) had performance levels differentiated by the rate variable only. When stimulus sources are close together, it makes no difference whether coherency is high or low—only whether rate is fast or slow. A Type I analysis for the four adjacent groups gave F ratios for Methods variance and the Trials by Methods variance that were significant beyond the .01 level of confidence. However the role of the rate variable is shown by performing the same Type I analysis separately for Groups HA-2 and LA-2, and for

[2] All subsequent references to a Type I design also will be for the final 10 trials with simultaneous time on target as the measure.

Groups LA-1 and HA-1. Neither of these lesser analyses, where the groups differ only in terms of stimulus coherency, allowed a rejection of the null hypothesis for the Methods variance, although their Trials by Methods variance was significant at the .01 level.

Separated sources.—When the stimulus sources are separated, both rate and coherency function to differentiate Groups HS-1, LS-1, HS-2, and LS-2. In contrast to adjacent sources where only rate made a difference, here we find that both rate and coherency contribute to between-groups differences. Apparently, just as with the demanding low-coherency condition, the wide separation of stimulus sources is exacting and makes a difference for all other variables. When the four separated groups are subjected to a Type I analysis, the Methods variance and the Trials by Methods variance permit rejection of the null hypothesis beyond the .01 level.

One-second rate.—A Type I analysis for Groups HA-1, HS-1, LA-1, and LS-1 produced a Methods variance and Trials by Methods variance significant beyond the .01 level. This significant difference among methods means was primarily contributed by the inferior performance of LS-1. A Type I analysis for only HA-1, HS-1, and LA-1 gave an F ratio for the Methods variance that lacked the .01 probability level, although the Trials by Methods variance was significant at the .01 level.

Two-second rate.—The ordering of 2-sec. groups is about the same as for 1-sec. groups. A Type I analysis for HA-2, HS-2, LA-2, and LS-2 gave an F ratio for the Methods variance which is significant beyond the .01 and a nonsignificant Trials by Methods variance. The methods difference was a function of the lower performance level of Group LS-2. Deletion of LS-2 from the Type I analysis resulted in a nonsignificant F value for methods.

All 2-sec. groups average a higher performance level than the 1-sec. groups. The reason for this would seem to be that response to each event can result in some slight off-target time and the 2-sec. groups had only half as many responses on a trial as the 1-sec. groups (see next section on Response Time Analysis).

Response Time Analysis

Proficiency in discrete tracking is primarily a function of the accuracy of timing the response to each stimulus change. To the extent that S is responding to a stimulus event as in a classical reaction time experiment where his response occurs after the onset of the stimulus, his off-target time will be equal to his reaction time which, in turn, is a function of the experimental conditions. However, if S can learn to anticipate certain classes of events under particular experimental conditions, then ideally his response time can be shorter than the expected reaction time value and his time-on-target score will be proportionately increased. Some anticipatory responses can be poor because they are so premature that they result in more off-target time than if S merely had waited for the stimulus event to occur and then responded as in classical reaction time. However, well-learned anticipatory responses will closely bracket the onset of the stimulus event and result in less off-target time than the optimum reaction time value (Poulton, 1952).

On Trial 50, for each S, each hand, each event, the time between the onset of the stimulus event and the

onset of the response was measured from the chart records of the recorder and analyzed to see how these response times vary as a function of the experimental treatments. If reaction time is taken to mean response latency to the occurrence of a stimulus, then it is perhaps best to refer to these measures as *response times* (RT) because anticipatory behavior results in some of them occurring coincident with the stimulus or before it, as will be shown. To reduce the thousands of RT measures to manageable proportions, an RT score for Trial 50 was computed for each *S*, each hand, white events, and colored events. Thus, each *S* had a Left-White RT Score, a Left-Colored RT Score, a Right-White RT Score, and a Right-Colored RT Score. Each *S*'s score was a mean of individual RT values for the hand and class of events indicated. If an RT preceded the onset of the stimulus it was given a positive sign, and if it followed the stimulus it had a negative sign. These signs were retained in the computation of the RT score. Complete omissions were separately categorized. An RT score involved only the response times of correct responses, which was virtually all responses on Trial 50. Red and green stimuli were pooled as a common category of uncertain colored events because their respective demands on *S* were essentially the same.

Table 2 gives the Trial 50 mean in milliseconds for each group for each of the RT scores. All of the values are negative, signifying that, on the average, response occurred after the onset of the stimulus. The row grand means are overall values which include all events for both hands. In interpreting RT values in terms of proficiency, it should be kept in mind that 2-sec. groups had half as many response changes as 1-sec. groups, low-coherency groups had fewer response changes than high-coherency groups because colored events were occasionally prolonged, and simultaneous time-on-target which, as time coincidence of response pairs, is only partly a function of average RT. Table 3 gives a summary of the results of a $2 \times 2 \times 2$ analysis of variance of each of the four Trial 50 RT scores.

A significant aspect of Table 2 is that the high-coherency, 1-sec. groups have impressively short RTs to white events, whether the stimulus sources are adjacent or separated.

TABLE 2

MEAN OF RT SCORES (MSEC.) FOR EACH GROUP ON TRIAL 50

Group	Left Hand		Right Hand		Grand Mean
	White Events	Colored Events	White Events	Colored Events	
HA-1	−185	−374	−181	−368	−277
HA-2	−373	−452	−406	−419	−412
LA-1	−444	−441	−434	−415	−433
LA-2	−450	−495	−475	−475	−474
HS-1	−175	−348	−181	−337	−260
HS-2	−454	−426	−372	−379	−408
LS-1	−600	−505	−566	−543	−553
LS-2	−660	−669	−673	−687	−672
Grand Mean	−418	−464	−411	−453	

TABLE 3

SIGNIFICANCE LEVELS FOR *F* TESTS OF RT SCORES

Source	Left Hand		Right Hand	
	White Events	Colored Events	White Events	Colored Events
Coherency (C)	<.01	<.01	<.01	<.01
Rate (R)	<.01	<.01	<.01	<.01
Separation (S)	<.01	<.01	<.01	<.01
C × R	<.01	>.05	<.01	>.05
C × S	<.01	<.01	<.01	<.01
R × S	>.05	>.05	>.05	>.05
C × R × S	>.05	>.05	>.05	>.05

TABLE 4

PERCENTAGE OF ALL RT VALUES ON TRIAL 50 IN THE RANGE OF ±133 MSEC. AND CONSIDERED BENEFICIALLY ANTICIPATORY

Group	Left Hand		Right Hand	
	Colored Events	White Events	Colored Events	White Events
HA-1	0.5 (0.0)	26.0 (8.0)	0.0 (0.0)	28.0 (5.0)
HA-2	0.0 (0.0)	1.0 (0.3)	0.0 (0.0)	1.0 (0.0)
LA-1	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.2 (0.0)
LA-2	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
HS-1	5.0 (1.1)	31.0 (7.2)	5.0 (0.3)	34.0 (7.0)
HS-2	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
LS-1	1.5 (0.9)	0.6 (0.0)	1.6 (0.3)	2.5 (0.9)
LS-2	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)

Note.—The values in parentheses are the percentage Trial 50 RT values that are perfectly anticipatory where stimulus and response coincided in time.

This might be expected on the basis of their event-time certainty alone, but the high-coherency, 2-sec. groups which have event-time certainty have white RTs which are two to three times longer. Short RT values are somehow a function of the duration of stimulus events, in addition to event-time certainty.

Table 4 elucidates this by showing that only Groups HA-1 and HS-1 have a sizable percentage of response times to white events that can be interpreted as beneficially anticipatory. This table gives, as the primary entry, the percentage of all RT measures made on Trial 50 that fall within the range ±133 msec. The scores determining these percentages are judged to be anticipatory. The rationale for this judgment is essentially the same as Poulton's (1952), namely, that if *S* waits for the stimulus to occur before responding, as in classical reaction time, he would have a minimum off-target RT in the approximate range of 150–250 msec. (Klemmer, 1956, 1957). The RT scores used for the percentages in Table 4 are substantially less than that and allow the reasonable inference that a positive anticipatory mechanism is operating. This is further supported by the values in parentheses which give the percentage of perfect responses where response occurrence exactly coincides with onset of the stimulus event and results in no off-target time whatsoever. Although the percentages of coincident responses tend to be small because of the stringent simultaneity requirement, Groups HA-1 and HS-1 have an impressively larger percentage of them than the other groups. Moreover, only these two groups had negative and positive RT values around zero, although the distribution was asymmetrical with the majority of them negative. All of the other groups had almost all negative values, with rare and random zero and positive RTs. It must be concluded that anticipatory behavior of the kind that enhances proficiency has high stimulus coherency and a fast rate among its determiners.

Figure 1 shows that the low-coherency, separated groups perform rather poorly, and Table 2 corroborates this by showing that these groups had relatively long RT values. Group LS-1 has particularly poor performance and Table 5 helps to

TABLE 5

NUMBER OF OMITTED RESPONSES ON TRIAL 50

Group	Left Hand		Right Hand	
	White Events	Colored Events	White Events	Colored Events
HA-1	0	0	0	0
HA-2	0	0	0	0
LA-1	0	0	0	0
LA-2	0	0	0	0
HS-1	0	2	1	1
HS-2	0	0	0	0
LS-1	53	11	50	16
LS-2	2	0	0	0

explain it further by showing a disproportionately large number of omitted responses for this group.

Discussion

Of basic significance for understanding complex tracking behavior is the role of stimulus coherency and its interrelations with the spatial separation of stimulus sources and the rate of change of stimulus events. When coherency is low, *S* is denied enough event regularities which make for prediction behavior so the response to each event in each source is much the same as in visual-motor reaction time studies where *S*'s response follows his direct sensing of it. Whether *S* is on a colored event and must prepare for a white event, or whether he is on a white event and has a colored event coming next, he is faced with uncertainty. When he is on a colored event there is time uncertainty because he can never be quite sure when the white event will occur. This constitutes a variable foreperiod, for reaction to white and Klemmer (1956, 1957) has shown that reaction time is increased under these circumstances. And, when *S* is on white, he is confronted with event uncertainty because red and green are equally probable next events. This also is known to increase reaction time (Bricker, 1955). As a result of these continual uncertainties *S* has a pressing requirement for visual scanning because he must always strive to keep oriented to each stimulus source as much as possible to know what response change is required of him. There is never enough certainty in the stimulus series for him to acquire a predictive capability which would reduce some of the visual scanning requirements. Visual scanning of two sources is rather easy when the stimulus sources are adjacent and event rate is slow, but the demands are substantially heightened when sources are separated and the rate of events increased. The expectation, which is supported by the data in Tables 2 and 5, is for missed or late responses because of the likelihood that *S*'s eyes and head will be momentarily oriented away from an event change. The low-coherency groups with the separated stimulus sources (Groups LS-1 and LS-2) had longer RT values (Table 2) than the two low-coherency groups with adjacent sources (Groups LA-1 and LA-2), and one of the reasons is surmised to be that *S*s often were not oriented to a source when an event change occurred and RT was increased by the delay of orienting to it. Table 5 shows that missed responses were a factor only for Group LS-1, the most difficult of the experimental conditions. The behavioral demands imposed by low coherency, fast rate, and wide separation of stimulus sources must, at times, be excessive and cause *S*s to completely omit responses.

The high-coherency series produced learned behavior which functions to reduce a number of the demands that the low-coherency series imposed on responding. It was of special interest to find that high-coherency events eliminate the pronounced difference between adjacent and separated sources found for the low-coherency series. There are two aspects of the behavior of high-coherency groups that need discussion and clarification. The first is why Groups HA-2 and HS-2 have equivalent terminal performance even though they do not have white RT values in the range considered beneficially anticipatory (Table 4). The simplest hypothesis is that the scanning response for Group HS-2 is faster than for Group HA-2 so that the sensing of events in two sources takes the same amount of time. If this is true it must hold only for high-coherency events because LS-2 has a distinctly lower performance than LA-2. A more elaborate hypothesis is that HS-2 is manifesting prediction of white events for one source while attending to the other but that prediction is not revealed in zero and very brief RT values as found for Groups HA-1 and HS-1. This is a tenuous explanation however, in the absence of any measured evidences of prediction behavior. Explanation of this finding must await new indices of anticipation.

The presence of prediction behavior for Groups HA-1 and HS-1 is the second intriguing aspect of the high-coherency series and is unequivocal because of the presence of zero and very brief white RT values that could not be expected had *S* responded after sensing the event change, as in classical reaction time. Here the challenging problem lies not so much with identifying predictive behavior as it does with asking the mechanism of prediction and how it is carried in the organism. There are two prominent kinds of prediction possibilities for visual-motor responding. The first has to do with an environmental stimulus complex that is changing dynamically in time and *S* has to learn to initiate a time-varying response to a particular stimulus configuration at Time t such that the response or an effect of it coincides with a later stimulus configuration at Time $t + 1$. Properly timing the swing of a bat to hit a pitched ball represents prediction of this sort. The batter must learn to start the swing early so the bat is in the correct position for a hit at the moment the ball crosses the plate. An explanation of this kind of behavior is quite amenable to S-R analysis, providing we remember that both the stimuli and responses are time-varying quantities. The *S* would appear to be starting the response to a present stimulus complex now at Time t and as the response continues in time it becomes, through learning, coincident with the environmental stimulus configuration at Future Time $t + 1$. Much of what we term skilled behavior embodies this kind of prediction.

The second type of prediction, which is most relevant for the task of the present experiment, concerns accurately timed responses where the present environmental stimulus complex is *constant*. Here we are confronted with a more difficult analysis because the unchanging external stimuli do not give distinctive differential cues to elicit a response change at the proper moment. An example is in the high-coherency task where the present colored event was steady and unchanging, thus providing no unique cues to signal *S* of the proper time to initiate his response change so that it will be in the new position coincident with the onset of the next white event. Apparently the cuing must arise within *S*. While some may be inclined to interpret these phenomena in terms of cognitive time perception, it is consistent with contemporary physiology to have internal time-keeping for motor responses be noncognitive and resident in proprioceptive neural networks. Consider that the movement to a colored event produces a large proprioceptive neural discharge. Because the task requirement is for *S* to hold the control in position as long as the event persists, no additional movement is made while the colored event is on and the initial proprioceptive discharge proceeds to decay in time. Presumably, with the extant conceptions of reverberating neural circuitry this decay can continue for a long time relative to the decay characteristics of isolated nerve fibers (Ruch, 1951). Regarding the proprioceptive discharge as internal stimuli, and knowing that it has the property of time decay, the hypothesis is tenable that we have a changing internal stimulus state and it is the unique configuration of this state at Time t which cues the initiation of a response movement that is coincident with a change in the external environment at Time $t + 1$. We would predict from this hypothesis that in addition to stimulus coherency, anticipatory behavior is a function of the time following the response producing the proprioceptive discharge and of control system parameters determining proprioceptive feedback. This study lends some support to the hypothesis by finding beneficial anticipatory responses only for high-coherency groups having 1-sec. events, not 2-sec. events.

Our findings are a helpful beginning in understanding the role of certain variables in two-dimensional visual tracking, but they also impress upon us our ignorance of prediction mechanisms and the observing response. Our knowledge of prediction is at an early empirical level and is insufficient for clarifying

the explanatory processes. Certain hypotheses have been suggested by the results but much research stands between them and verification. A good deal of this research could be carried out profitably on simpler tasks. In contrast, the observing response is most prominent in multidimensional tasks and is best studied under these more complex conditions. The evidence for the observing response in this study has been inferential and did not involve measurement of eye and head movements. This approach of direct measurement is obviously important, not only for knowing what is seen when, but in determining the conditions (if any) under which two spatially separated events can receive differential response without scanning eye movements. Recent technical advances in the measurement of eye movements by a closed-circuit television technique (Mackworth & Mackworth, 1958) show promise as a means of accurately correlating eye movements and stimulus events.

Summary

A two-dimensional discrete pursuit tracking apparatus with two visual stimulus sources and a control stick for each hand was used to investigate two values each of stimulus coherency, spatial separation of stimulus sources, and the rate of change of events. Stimulus coherency was defined in terms of the probabilistic organization of the three stimulus lights in each source, and the two values used were high and low coherency. The rate of event change was 1 and 2 sec., and spatial separation was 5° and 30° on each side of the median plane of *S*'s visual field. It was conjectured that *S*'s acquired ability to predict certain events in the high-coherency case would positively influence tracking proficiency by permitting *S* to respond to a source in the absence of direct visual attention, thereby reducing the requirement for visually scanning the sources and responding to all events by direct visual sensing. The rate of event change was expected to be interactive with the other two variables because it is a determiner of the observing response rate. The experimental design was a $2 \times 2 \times 2$ randomized factorial arrangement of the eight conditions, with 12 *Ss* in a group. Fifty 100-sec. trials were administered over three testing sessions. Intertrial rest interval was 40 sec. The principal measure of performance on a trial was the time that *S* was simultaneously on target in both sources.

The findings revealed that tracking proficiency was related positively to the proximity of stimulus sources and inversely to the speed of event change when stimulus coherency was low. But, when coherency was high, only event rate was a differentiating variable. Spatial separation did not distinguish between the groups of any given rate condition, presumably because *S* was utilizing his capability for predicting high-coherency events to reduce the demands of visual scanning when stimulus sources were widely separated. Results were discussed in terms of hypotheses for prediction behavior in visual-motor responding.

References

Adams, J. A. Human tracking behavior. *Psychol. Bull.*, in press.

Attneave, F. *Applications of information theory to psychology.* New York: Holt, 1959.

Bartlett, F. C. Anticipation in human performance. In, *Essays in psychology: Dedicated to David Katz.* Upsala: Almquist & Wiksells, 1951. Pp. 1–17.

Bricker, P. D. Information measurement and reaction time: A review. In H. Quastler (Ed.), *Information theory in psychology.* Glencoe, Ill.: Free Press, 1955.

Charipper, B. A. Ship control. VIII: Single element vs. two-element display in two-dimensional tracking. *Electric Boat tech. Rep.*, 1959, No. SPD 59-003.

Fitts, P. M., & Simon, C. W. Some relations between stimulus patterns and performance in a continuous dual-pursuit task. *J. exp. Psychol.*, 1952, **43**, 428–436.

Helson, H. Design of equipment and optimal human operation. *Amer. J. Psychol.*, 1949, **62**, 473–497.

Jackson, K. F. Behaviour in controlling a combination of systems. *Ergonomics*, 1958, **2**, 52–62.

Klemmer, E. T. Time uncertainty in simple reaction time. *J. exp. Psychol.*, 1956, **51**, 179–184.

Klemmer, E. T. Simple reaction time as a function of time uncertainty. *J. exp. Psychol.*, 1957, **54**, 195–200.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

Lockard, R. B. Bibliography of human engineering reports on tracking. *USN Ord. Test Station Rep.*, 1956, No. 5272.

MACKWORTH, J. F., & MACKWORTH, N. H. Eye fixations recorded on changing visual scenes by the television eye-marker. *J. Opt. Soc. Amer.*, 1958, **48**, 439–445.

POULTON, E. C. Perceptual anticipation in tracking with two-pointer and one-pointer displays. *Brit. J. Psychol.*, 1952, **43**, 222–229.

POULTON, E. C. On prediction in skilled movements. *Psychol. Bull.*, 1957, **54**, 467–478.

RUCH, T. C. Motor systems. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 154–208.

SENDERS, V. L. The effect of number of dials on qualitative reading of a multiple dial panel. *USAF WADC tech. Rep.*, 1952, No. 52–182.

SWEENEY, J. S., BAILEY, A. W., & DOWD, J. F. Comparative evaluation of three approaches to helicopter instrumentation for hovering flight. *USN Res. Lab. Rep.*, 1957, No. 4954.

WYCKOFF, L. B. The role of observing responses in discrimination learning. Part I. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received November 16, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 404–407

INDIVIDUAL DIFFERENCES IN LEARNING AS A FUNCTION OF SHOCK LEVEL

RACHEL KAPLAN, STEPHEN KAPLAN, AND EDWARD L. WALKER

University of Michigan

It is a well known fact that as the level of motivation is increased the quality of learning tends first to increase and then to decrease as motive levels reach high values. Although theoretical explanations of improved performance with increases in motivation differ rather widely, the phenomenon itself does not seem inherently strange. On the other hand, the fact that performance breaks down at high levels of motivation does seem inherently curious. The present study is a first step in the exploration of mechanisms of decrements in performance at high levels of motivation.

One of the major difficulties in research on this problem stems from the fact that curves relating performance to some measure of motivation are usually curves of group performance. Such curves often misrepresent the relation between motive level and performance in a single individual. Ideally, one would want a curve representing the performance of the same *S* at a number of motive levels. Such curves are often difficult to interpret because they involve repeated testing of the same *S* and thus entail an interaction between successive tests. Performance at any one value of the motivating operation is then confounded by the effects of tests at other levels. The present study is a compromise between an analysis of group performance and the analysis of individual performances.

A long series of preliminary studies confirmed the observation that at a given shock level on the grid floor of a T maze there are wide individual differences in the behavior of the animals. Three classes of behavior can be recognized easily and almost unmistakably. They are: (*a*) *unmotivated behavior* in which *S*s seem not to mind the shock and not to learn, (*b*) *learning behavior* in which *S*s learn and can readily reverse sides if the escape side is changed, and (*c*) *fixated behavior* in which *S* may or may not learn to choose the correct side but tends to remain inflexibly with his first choice regardless of which side permits escape from the shock.

While all three types of behavior might be seen at certain levels of shock in the maze, the "unmotivated" kind is clearly more evident at low levels of the shock, and the "fixated" kind is more frequently observed at higher levels. Furthermore, during pilot studies the observation was made that *S*s displaying each type of behavior in the maze seemed to display different degrees of "emotional" behavior while being handled during the experiment. If *S* was difficult to place in the maze and to remove from it, if it squeaked frequently and defecated frequently, it seemed more likely to display "fixated" behavior. If it was very easy to handle, was quiet, and did not defecate while being handled, it was more likely to show "unmotivated" behavior in the maze.

With these observations as a basis, an experiment was designed which permitted the investigation of the frequency of each type of behavior in the maze as a function of the intensity of shock on the floor. It

also permitted an exploratory investigation of the relation of a crude measure of individual differences in "emotionality" to various types of behavior in the maze.

METHOD

Subjects

The *S*s were 68 male albino rats ranging in age from 90 to 150 days at the time of the experiment. They were experimentally naive.

Apparatus

The three alleys of a simple enclosed T maze were approximately 18 in. in length and approximately 4 in. wide and 4½ in. high. The floor of the maze was a shock grid throughout. The maze was covered with transparent plastic over the choice arms and with hardware cloth over the stem and goalboxes. There were three guillotine gates, one permitting release from the startbox and the others permitting confinement to the goalboxes. The colors of the two choice arms could be controlled by exchanging black and white inserts. The experiment was carried out in a small, windowless room jammed with other apparatus. The maze was illuminated by a 150-w. bulb 2 ft. above it.

The shock employed was a dc square wave at 120 pulses/sec. with a 2-msec. pulse duration. There was no appreciable resistance other than *S* in the circuit. Six shock levels were used: 20, 25, 30, 35, 40, and 45 v. A milliammeter wired in series with the stimulator informed *E* when the grid was shorted out. A clock attached to the stimulator measured latencies.

Procedure

Each *S* received 40 consecutive trials using the correction procedure. All trials for a given *S* were run at a given shock level in a single session with trials following each other as rapidly as unhurried handling of *S*s would permit. There were 12 *S*s at each of the five higher shock levels, and 8 *S*s at the lowest level, 20 v.

The detailed handling procedure was as follows. On a day preceding the experimental day each *S* was handled for a 10-min. period. On the experimental day *S*'s feet were dipped in a saline solution and then blotted on a towel. This procedure was used to minimize individual differences in footpad moisture. The *S* was then placed in the startbox of the maze. The startdoor was lifted and the shock was turned on simultaneously. The shock was on continuously for 60 sec. unless *S* reached the correct goalbox whereupon the door was closed behind it and the shock turned off. On a correct choice *S* remained in the goalbox 15 sec. before removal. If *S* failed to reach the correct goalbox in 60 sec., it was removed from the maze. Each subsequent trial started with the dipping of the feet in the saline solution, and the procedure was repeated 40 times.

The *S*s were tested in blocks of four. Either the black side or the white side was initially correct and the positive side could be either on the right or on the left. Each block of four included one *S* started under each of these four conditions. When, and if, *S* correctly chose the designated correct side at least 6 out of 8 trials with at least the last 3 consecutive choices correct, it was said to have met the criterion of learning, and the other side became correct. The position of the colored inserts remained constant for an individual *S*. It was possible for *S* to meet the criterion of learning a maximum of five times during 40 trials.

A set of rating scales of "emotionality" was developed and applied during the first 20 trials. Each *S* was rated on a 4-point scale (0–3) for each of four measures: ease of handling when being placed in the maze, ease of handling during removal, amount of vocalization, and amount of defecation. Theoretically, a score of 0 to 12 could result. It cannot be argued that such "emotionality" ratings were truly independent of the maze performance, for the ratings were made by *E* during the runs. However, an effort was made by *E* to make them as objectively as possible, and they do have a degree of value as preliminary or exploratory data.

Performance Criteria

Unmotivated performance.—An *S* was included in this class if it never met the criterion of learning and if there was variability in the choice of alternatives during the 40 trial session.

Learning performance.—An *S* was classified as having learned if it met the criterion of learning at least once and showed flexibility in its behavior. All but one *S* included in this category demonstrated flexibility after meeting the criterion for the first time. The exception was a 25-v. *S* which met the initial learning criterion on Trial 34. This *S* showed considerable variability in its be-

havior in meeting the criterion for the first time and was thus clearly an *S* manifesting "learning" behavior.

Fixated performance.—An *S* was classified in this group if it took all of its runs to the "wrong" side, or if it met the criterion of learning in 12 or fewer trials and then continued in that choice for the remainder of the 40 trials. The only *S* included in this group showing any flexibility in performance met the criterion in the first 9 trials, repeated its response for 25 more trials and changed to the now correct side in the last 6 trials.

RESULTS

The primary results may be seen in Fig. 1. The upper curve represents a plot of the expectations based on the earlier pilot results. The expected number of "unmotivated" *S*s decreases rapidly as shock level is increased. It was expected that the number of *S*s exhibiting "learning" behavior would increase and then decrease. Finally, the number of *S*s showing "fixated" behavior was expected to increase to include all *S*s tested at very high shock levels.

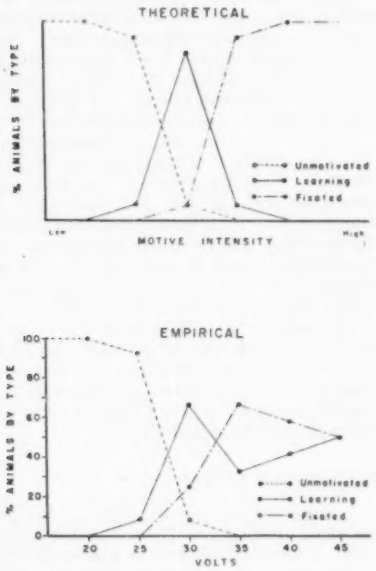

FIG. 1. The upper figure indicates the general expectancy with respect to the numbers of *S*s which would show "unmotivated," "learning," or "fixated" behavior as shock level is increased. The lower figure is a plot of the observed frequencies of each type of behavior. (There are 12 *S*s at each voltage level except that there are 8 *S*s at 20 v. See the text for the criteria for each type of behavior.)

The lower set of curves present the empirical results. It can be seen that the number of "unmotivated" *S*s decreased as expected. The number of "learning" *S*s increased and then decreased as expected, but then, contrary to expectation, began to rise again. This same failure of expectations also appeared in the curve of "fixated" *S*s. It rose, as expected, and then began to fall.

Observation of the behavior of *S*s at high shock levels reveals a possible source of the increase in number of *S*s meeting the learning criterion at 40 and 45 v. as compared to the number meeting the criterion at 30 and 35 v. At the higher shock levels an increasing number of apparently "fixated" *S*s were unable to cling to the grid for the full 60 sec. before removal from the maze. These *S*s bounced around in the maze and eventually happened to reach the correct side, thus meeting the objective criterion of learning, even though their behavior gave no evidence of the plastic character of the behavior of *S*s that learned. Although they met the objective criterion of learning, it is doubtful whether they were responding to the relevant black and white stimuli of the goal arms.

An analysis of the individual difference measure of "emotionality" is shown in Table 1. The mean rating increases with the increase in shock level in a fairly orderly manner. More important, however, are the comparisons between *S*s showing dif-

TABLE 1

"EMOTIONALITY" SCORE IN RELATION TO SHOCK LEVEL AND MAZE PERFORMANCE

Shock Level in Volts	Maze Performance Type					
	Unmotivated Behavior		Learning Behavior		Fixated Behavior	
	N	Mean	*N*	Mean	*N*	Mean
20	8	2.75				
25	11	3.09	1	2.00		
30	1	4.00	8	4.50	3	7.00
35			4	6.75	8	7.63
40			5	7.00	7	7.57
45			6	7.16	6	9.00

ferent maze behaviors at the same shock level. In every case the mean "emotionality" score for *S*s showing "fixated" behavior is greater than that of *S*s showing "learning" behavior at the same shock level. An analysis of variance of these results indicates that there is a statistically significant relation between "emotionality" score and shock level and between "emotionality" score and type of maze behavior, both of which are significant beyond the .001 level. Furthermore, detailed analysis shows that there are significant differences between "emotionality" scores of *S*s displaying "learning" behavior as against *S*s displaying "fixated" behavior which is significant beyond the .025 level at 30 v. and at the .05 level at 45 v.

If the "emotionality" index can be taken as a crude measure of individual differences in sensitivity to shock, the relations between "emotionality" and performance within shock levels would tend to indicate that if shock strength could be scaled in psychological units, there would be even less overlap in maze performance types than is apparent in Fig. 1. There, performance is matched against a physical scale of shock strength. It should be noted that individual differences in emotionality may be correlated with the resistance the rat contributes to the circuit. If this were the case, the behavior of the emotional *S*s might be explained in terms of the greater current they received. This is a matter for future investigation.

It can be concluded from these results that the rise and fall in efficiency of learning often observed to accompany progressive increases in motive strength, especially when motive strength is manipulated by variation in shock strength, may actually be composed of qualitatively as well as quantitatively different kinds of behavior. It is suggested that the breakdown in performance at high motive levels may actually be an instance of highly efficient learning of a task other than that set by *E*. In this instance it is suggested that the high shock levels may induce such devoted attention to the feet that the animal does not "notice," and thus can not respond to, the visual cues provided by *E*.

SUMMARY

Six groups of rats were tested with the correction procedure in a simple T maze and with either 20, 25, 30, 35, 40, or 45 v. of shock (dc square wave with 120 pulses/sec. and 2-msec. duration/pulse) on the entire floor of the maze. Training consisted of 40 massed trials in a single day. When *S* met the learning criterion of 6 out of 8 choices to the correct alternative the "correct" side was reversed. It was predicted that as the shock level was increased, the number of *S*s showing "unmotivated" behavior would decrease, the number showing "learning" behavior would increase and then decrease, and that the number showing "fixated" behavior would increase. This expectation was realized except that at the two highest shock levels an increasing number of *S*s met the objective criterion of learning. A post hoc interpretation of this particular result is given. Individual differences in "emotionality" bore a significant relationship to type of maze performance within a shock level, indicating that a psychological scale of shock intensity might have distinguished the three types of maze performance even more sharply than the physical scale of voltage.

(Received November 18, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 408–415

COMPARISON OF VERBAL RESPONSE TRANSFER MEDIATED BY MEANINGFULLY SIMILAR AND ASSOCIATED STIMULI [1]

JAMES J. RYAN

University of Nevada

In studies concerned with the process of transfer in learning, one of the factors usually considered crucial in determining the amount of transfer has been the specific relation existing between the original and subsequent stimulus or response elements. The degree of *similarity* between these elements, with respect to any of several attributes, has characteristically been represented as the basis for predicting and explaining differences in transfer. With respect to verbal materials, several studies (Haagen, 1943; Osgood, 1946, 1948; Underwood, 1951) have provided empirical support for similarity as a factor contributing to transfer. But, except for variation in the degree of similarity between subsequent stimulus and response terms in paired-associate learning, little work has been done investigating the other properties of the relationship between these verbal materials which might also facilitate transfer. However, some studies (Bulgelski & Scharlock, 1952; McClelland & Heath, 1943) have indicated that verbal habit connections existing between interlist stimulus or response words can mediate effects not unlike those which contribute to transfer effects.

Elaboration of the similarity concept in verbal materials has culminated in a theory presented by Osgood (1953) which hypothesizes that similarity in meaning has a dominant role in the mediation of verbal learning. Osgood suggests that covert meaning reactions can function as mediators in this process and also has suggested that observed positive and negative transfer effects could be explained in terms of differential mediating effects accompanying differences in meaningful similarity. In addition, Osgood and Suci (1952) have developed a scaling procedure which provides a basis for deriving a measure of the degree of connotative similarity between words.

In recent years, under the impetus of concern with communication and the psychology of language, interest has developed in intraverbal habit chains manifested by frequency of association in free-association response norms. Research investigating effects of such verbal habits on behaviors such as recall, perception, learning, and problem solving (Jenkins & Russell, 1956, 1958; Russell & Storms, 1955) has led to consideration of these pre-existing verbal habits as a basis for mediation of transfer in the traditional retroactive inhibition or response transfer paradigms. Bastian (1956) recently conducted a study directly concerned with the comparison of meaningful similarity and verbal habit connections as mediators of interlist transfer. In his study, verbal responses paired with the

[1] This article is based upon a dissertation submitted to the University of Minnesota in partial fulfillment of the requirements for the PhD degree. The author is indebted to James J. Jenkins for his helpful direction and generous advice on all phases of this study.

same stimulus terms in sequential paired-associate learning lists represented the different interlist relations being compared. He obtained significantly greater interlist transfer effects for response words having strong verbal habit connections and minimal similarity in meaning as compared with response words having high similarity in meaning and weak verbal habit connections.

The present study represents an extension of this comparison to a paradigm in which the interlist *stimulus* words are differentially related with respect to meaningful similarity and verbal habit strength. The purpose of this study was to determine the relative effectiveness of each of these verbal factors for mediating response transfer between stimulus words in two successively learned lists of paired associates and in the subsequent recall and relearning of the first list. To provide equivalence between the magnitudes of the two factors being compared, the objective was to determine the contribution of each factor independently when it is at maximum strength and other factors are of relatively minimal strength. This study differs from most previous studies concerned with similarity in meaning in that the semantic differential served as the basis for determining the degree of similarity in connotative meaning in terms of derived semantic distances (D). As in Bastian's study, free association served as the basis for inferring the pre-experimental strength of the verbal habit connections.

METHOD

Experimental Design

The experiment was designed so that one basic list of words would serve as stimulus terms in either the original list (Sequence I condition) or the transfer list (Sequence II condition). For each of these basic-list stimulus words, three other words were selected to represent one of the following types of relationship with respect to the basic-list stimulus words:

Type A (associated)—*Maximal* habit strength connections with *low* similarity in meaning.

Type S (similar)—*Weak* habit strength connections with *maximal* similarity in meaning.

Type C (control)—*Weak* habit strength connections with *low* similarity in meaning.

The necessity of having the same response paired with each of four different stimulus words required three separate paired-associate lists in addition to the basic list. Each *S* was required to learn the single set of responses to only two paired-associate lists, the basic list and one of the other three lists. The latter were made up in a counterbalanced order so that each list contained an equal number of stimulus words representing each type of relationship when the list was learned in sequence with the basic-list stimulus words. Each *S* was then exposed to an equal number of stimulus words in the transfer list representing each relationship, thereby allowing an estimate of the major experimental differences independent of subject differences.

There were six conditions in all, three in which the basic-list stimulus words occurred in the original list with *one* of the other three sets of stimulus words occurring in the transfer list (Sequence I); and three in which the basic-list stimulus words occurred in the transfer list and the other set of stimulus words occurred in the original list (Sequence II).

Selection of Materials

Stimulus words.—The indices most closely indentified with the two processes being compared served as the primary basis for selecting stimulus words representing the pre-experimental verbal relationships. Two normative studies, using a student sample from a population very similar to that from which the experimental *Ss* were drawn, provided the data for these indices.

Russell and Jenkins (1954) have obtained free-association response norms for the 100 Kent-Rosanoff stimulus words. These norms indicate the primary (most frequent) response given to each stimulus word. A maximal habit strength connection is assumed to

TABLE 1

WORD RELATIONSHIPS AND LIST ASSIGNMENTS

Basic List	Stimulus Words						Response Words
	Type A		Type S		Type C		
	Words	List	Words	List	Words	List	
SLOW	FAST	3	DIM	1	FLAMING	2	CORLEP
SQUARE	ROUND	2	LONG	3	CURLED	1	TIFTON
HARD	SOFT	1	RIGID	3	MILD	2	BURBET
HEALTH	SICKNESS	3	VISION	2	ANGER	1	MERMIC
WAR	PEACE	1	DANGER	2	CALM	3	LODRAM
SWEET	BITTER	2	RIPE	1	CRUDE	3	VOLVAP
BOY	GIRL	2	FATHER	3	SISTER	1	DUNTEL
UGLY	BEAUTIFUL	3	HEARTLESS	1	CHARMING	2	WENDAN
FOOD	HUNGRY	1	CLEAN	2	WET	3	RISPUL

exist between a given word and its primary response. Jenkins, Russell, and Suci (1957) have made available a table of semantic distances (D values) between all pairs of a set of 360 words. This set includes a large number of Kent-Rosanoff stimulus words and also their primary responses as given in the Russell-Jenkins norms. The D value between a pair of words is derived from their individual semantic differential profiles and according to Osgood represents their relative degree of similarity in connotative meaning. From the set of words having data available on both indices it was possible to select the nine basic-list stimulus words such that for each of these there were three other words in the set which represented each of the desired interlist stimulus relationships. The stimulus words used are shown on Table 1.

In terms of the indices used these interlist relationships were as follows: (*a*) Type A words, for which the basic-list stimulus words were primary free-association responses and which had relatively high semantic distance values (low similarity of meaning) with respect to the basic stimulus words. (*b*) Type S words, for which the basic-list stimulus words had the lowest semantic distance values (highest similarity of meaning) and which had very low or zero frequency of occurrence as free-association responses with respect to the basic stimulus words. (*c*) Type C words, for which the basic-list stimulus words had a very low or zero frequency of occurrence as free-association responses and which had relatively high semantic distance values (low similarity of meaning) with respect to the basic stimulus words.

Each Type C word was selected to have approximately the same D value (low similarity of meaning) relative to the basic-list stimulus word as the corresponding Type A word while having the same relatively minimal degree of association strength as the corresponding Type S word. Stimulus words were also screened to include only those having high frequency of usage as indicated by the Thorndike-Lorge word count (Thorndike & Lorge, 1944); to minimize pre-experimental associative connections between Types A, S, and C words paired with the same response; and to minimize intralist interference effects.[2]

Response words.—To reduce the possibility of any associative or meaningful connections which would contribute to consistent but differential learning of a given pair or pairs of words in these lists, a set of two-syllable nonsense words (paralogs) was prepared for use as the response terms. This set of paralogs was pretested for consistent association value by including them in a free-association test and eliminating those that elicited a given response with any degree of consistency. These words were also constructed to minimize interlist and intralist orthographic similarity. The response words are shown in Table 1.

Procedure

A total of 90 male students randomly selected from sophomore general psychology courses served as *Ss*. The apparatus used was a Hull-type memory drum set up for

[2] A complete description of the stimulus words in terms of the relevant measures has been presented by the author (Ryan, 1957).

paired-associate learning. The apparatus presented each stimulus word for 2 sec. after which a shutter lifted to expose the response word (together with the stimulus word) for 2 sec. before a new stimulus word appeared alone. Stimulus materials were presented through use of tapes which contained randomly ordered repetitions of stimulus and response pairs in each list. Each *S* was told that he was to engage in a learning experiment and instructed in the function of the apparatus and the specific learning procedure to be followed. The *Ss* were instructed to say each stimulus word aloud as it appeared, and as soon as possible after the first trial to anticipate and say aloud each response before it appeared. If no response or an incorrect response was given, *Ss* were to say the correct response aloud when it did appear. The *Ss* were also told that each list would be continued until all responses had been correctly anticipated on the same trial.

As a warm-up task *Ss* learned a practice list consisting of five pairs of three-letter nonsense syllables as stimuli and two-digit numbers as responses. Following the warm-up task each *S* was given a set of 45 cards each containing one of the nine paralog response words. He was asked to go through the cards twice pronouncing each word aloud, thereby providing 10 exposures to each of the response words. This procedure allowed *Ss* to learn more readily the verbal response to the visual stimuli provided by these unfamiliar pafalogs eliminating somewhat this additional learning process from the actual experimental learning task. The *S* was then randomly assigned to one of the six experimental conditions and the experimental learning task was begun with the original list followed by the transfer list and relearning of the original list. Prior to beginning the original list, instructions were repeated and *S* was instructed not to attempt to anticipate the correct response on the first trial but only to say the response after it appeared. Each list was continued with a 4-sec. interval between each trial. Immediately after the original list had been learned to the criterion, *S* was instructed that a new list would be learned following the same procedure. The *S* was told that there would be a different set of stimulus words with the same set of response words. The learning of the transfer list proceeded without further delay. After the transfer list had been learned to the criterion, *S* was told that the original list would be repeated and that he was to attempt to anticipate the correct responses prior to their exposure for the first as well as subsequent trials.

Results and Discussion

The primary comparison in this study concerns response acquisition during transfer-list learning as well as performance during recall and relearning of the original list. However, this comparison assumes that there are no initial systematic differences in ease of response acquisition between those sets of stimulus words representing Types A, S, and C interlist relations as they appear in the transfer and relearning lists. To check this assumption a comparison was made of performance during the original-list learning between stimulus words representing each of these relationships. The reversal of sequence of presentation permitted this comparison. Stimulus words in the original list of each sequence could be classified on the basis of the functional relationship they represented in the transfer list of the alternate sequence.

An analysis of variance using the total correct anticipations given during original-list learning for Sequences I and II separately revealed no significant differences ($F = .18$ and .95, respectively, $P > .05$) in ease of response acquisition between words representing the three types of relationship.[3] This result is consistent with the assumption that stimulus words representing the three types of relationship do not differ initially with respect to ease of response acquisition.

[3] Since measures of the three experimental word types were obtained from the same *Ss* and were therefore intercorrelated within list conditions, an analysis of variance model presented by Collier for "Two dimensional classification with observations at one level correlated" (Collier, 1956) was used for this and all other analyses of word-type effects.

The performance measures which served as indices of the degree of transfer obtained during learning of the transfer list were the number of correct responses given on Trial 2 (first anticipation trial); the number of correct responses given on Trial 3; and the total number of correct responses given on all trials to the criterion. For each of these measures an analysis of variance was carried out for Sequences I and II separately. Each analysis of variance indicated a significant difference at the .01 level between word-type means for all measures with the exception of the total correct responses in Sequence II for which the probability was at the .05 level. Duncan's multiple range test (Duncan, 1955) was used to determine the significance of the difference between each pair of the three word-type means in each sequence. The means and *SD*s obtained for each measure are shown in Table 2.

Words representing the Type A relationship obtained a significantly greater number of correct responses on the initial anticipation trials (Trials 2 and 3) than words representing either Type S or Type C relationships. The differences between Type A and Type C words are significant at the .01 level in both sequences on Trials 2 and 3. The differences between Type A and Type S words are significant at the .01 and .05 levels, respectively, for Trials 2 and 3 in both sequences. A difference favoring the Type A words is also observed when the total number of learning trials are considered, although for this measure only the difference between Type A and Type C words is significant and this difference is at the .01 level in both sequences. The Type S words also exhibited a consistently greater ease of response acquisition than the Type C words for all measures of learning. This difference, though, is not significant for the first anticipation trial (Trial 2) of either sequence nor for subsequent trials in Sequence II. In Sequence I the difference between Type S and Type C words is significant at the .05 level for Trial 3 and at the .01 level for all trials to the criterion. These results indicate that the interlist relations between stimulus words represented by both the Type A words (strong verbal habit connections) and the Type S words (high similarity in meaning) facilitated learning to a greater extent than those for the control words. The Type A words however, exhibited greater ease of response acquisition than the Type S words as indicated by differences obtained on the initial trials. It appears then that the Type A relationship provided an initial learning advantage which carried over the subsequent trials and although the Type S relationship did not provide a reliable difference on the initial

TABLE 2

TRANSFER LIST LEARNING: MEAN CORRECT ANTICIPATIONS AND *SD*s

Sequence	Types					
	A		S		C	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
	Trial 2					
I	1.38	.89	.98	.69	.80	.81
II	1.53	.84	1.04	.90	.93	.81
	Trial 3					
I	1.93	.86	1.56	.87	1.24	.93
II	1.91	.92	1.56	.99	1.47	1.01
	All Trials					
I	22.60	12.26	21.36	11.47	17.91	7.67
II	22.02	10.53	20.53	10.49	19.07	10.77

trials, it did facilitate subsequent learning at a faster rate than was obtained for the Type C words as indicated by an overall difference with respect to the latter.

An analysis was made of performance during relearning of the original list. The results of the analysis of variance of the number of correct responses on the first relearning trial, which represents a measure of recall, indicated no significant differences between word-type means. Similarly the analysis of variance for the total number of correct responses given during all trials for relearning the original list indicated no significant effect attributable to word types. The means for both measures during relearning the original list are shown in Table 3.

Since the maximum number of correct responses possible for *S* to make on a single trial for a given word type was three, it can be seen that on the recall trial there was a very high probability of correct anticipation for words representing all three interlist stimulus relationships. It is apparent that there was a minimal loss in learning due to the interpolated transfer list and that this minor loss showed no systematic relationship to the different verbal relations.

TABLE 3

RELEARNING ORIGINAL LIST: MEAN CORRECT ANTICIPATIONS AND *SD*s

Sequence	Types					
	A		S		C	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
	Trial 1 (Recall)					
I	2.36	.84	2.22	.79	2.29	.82
II	2.31	.67	2.31	.70	2.16	.85
	All Trials					
I	7.96	4.60	7.91	4.97	7.58	4.20
II	9.49	7.83	9.78	7.74	8.96	6.11

On the basis of these results it can be concluded that the mediation effects of verbal habit strength (as determined by free-association response norms) and similarity in meaning (as determined by the semantic differential) can facilitate response transfer between interlist stimulus words. Further, for the sample of words used here, the mediation effects of the verbal habit connections provide greater facilitation than those provided by similarity in connotative meaning.

Since transfer effects have been obtained for words representing both verbal habit and similarity relationships, it is somewhat difficult to make any unequivocal statement favoring only one of the underlying theoretical positions. The response chaining mediation mechanism supported by Bastian (1956) and the Minnesota investigators and the representational mediating process offered by Osgood (1953) both find adequate support in terms of the transfer effected by the verbal relationships assumed to provide maximum opportunity for each of these processes to be effective. The most conservative conclusion that could be drawn at this point is that both factors, and as far as can be determined their assumed mediating processes, operate independently—each factor being a sufficient condition in itself for facilitating response transfer in a learning scheme such as that utilized here. It should be understood, however, that these results provide no indication concerning relative transfer effects of dimensions of similarity other than connotative similarity that could exist between stimulus words either separately or in combination with similarity in connotative meaning.

Although no systematic bias was apparent, the words used in this study to represent the interlist relations were not randomly sampled from the potentially large population of word pairs that could exhibit these relations, but

rather were selected from those for which the appropriate normative data were available. Therefore, more extensive generalization of these results concerning the relative effectiveness of the two factors would require some qualification.

It is consistently evident, however, that the transfer obtained through the verbal habit relationships used in this study is of larger magnitude than that obtained for the similarity relationships employed here. It might be suggested that this difference has arisen through a deficiency in the sampling procedures as far as similarity is concerned. That is, it is reasonable to question whether those words selected for maximal similarity (low semantic distance) to the basic-list stimulus words did not actually exhibit as much similarity as other possible words not so selected. For example, synonyms in general usage tend to have an association frequency of measurable magnitude and therefore, even if within the set for which normative data were available, could not be included here to represent the Type S relationship. Although it is possible that the degree of similarity may have been depressed to some extent through this kind of selection it should be noted that the semantic distances for the meaningfully similar words used here tended to be only slightly larger than the test–retest difference in semantic distance of words measured by the semantic differential obtained for a separate subject sample (Norman, 1957). This suggests that the error of measurement could account for any slight differences that would separate the words used from others that might appear more similar with respect to connotative meaning.

The results in this study seem to indicate quite clearly that verbal habit strength can mediate verbal response transfer independent of the effects of connotative similarity. Further, it seems that for certain subsets of words at least, verbal habit strength provides as good if not a better basis for verbal response transfer than is provided by connotative similarity in meaning.

Summary

Similarity in connotative meaning and verbal habit strength as separate factors relating interlist stimulus words in subsequent paired-associate learning lists were compared as to their effectiveness in mediating response transfer across these lists. The *S*s learned in sequence two paired-associate lists, an original and transfer list, and relearned the original list. Three types of verbal relationship were established between the corresponding interlist stimulus words paired for learning with the same response: Type A, representing maximal habit strength connections and low similarity in meaning; Type S, representing weak habit strength connections and maximal similarity in meaning; Type C (control), representing weak habit strength connections and low similarity in meaning. These interlist relations were obtained by pre-experimental selection of stimulus word pairs on the basis of appropriate norms for semantic distance (as derived from Osgood's semantic differential), which served as an index of similarity in connotative meaning; and free-association response frequency, which served as an index of verbal habit strength. Measures of performance on the transfer list showed that both Type A words and Type S words facilitated learning to a greater extent than the Type C control words and that the Type A words facilitated learning to a greater extent than the Type S words. No systematic differences in recall or relearning of the original list were obtained. It was concluded that both verbal habit strength and connotative similarity in meaning as relations between interlist stimulus words contribute independently to the mediation of response transfer and that the mediation effects contributed by verbal habit strength connections were as good if not better than those contributed by connotative similarity within the subset of words used here. The theoretical relevance of these two factors was presented and discussed.

References

Bastian, J. R. Response chaining in verbal transfer. Unpublished doctoral dissertation, Univer. Minnesota, 1956.

Bulgelski, B. R., & Scharlock, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334–338.

Collier, R. O., Jr. Experimental designs in which observations are assumed to be

correlated. Unpublished doctoral dissertation, Univer. Minnesota, 1956.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

HAAGEN, C. H. Learning and retention as a function of the synonymity of original and interpolated tasks. Unpublished doctoral dissertation, State Univer. Iowa, 1943.

JENKINS, J. J., & RUSSELL, W. A. Research studies in the role of language in behavior. *ONR tech. Rep.*, 1956. (Contract N8-onr-66216, Univer. Minnesota.)

JENKINS, J. J., & RUSSELL, W. A. The role of language in behavior. *ONR tech. Rep.*, 1958. (Contract N8-onr-66216, Univer. Minnesota.)

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. Studies on the role of language in behavior: An atlas of semantic profiles for 360 words. *ONR tech. Rep.*, 1957. (Contract N8-onr-66216, Univer. Minnesota.)

MCCLELLAND, D. C., & HEATH, R. M. Retroactive inhibition as a function of degree of association between original and interpolated activities. *J. exp. Psychol.*, 1943, **33**, 420.

NORMAN, W. T. The role of language in behavior: Stability characteristics of the semantic differential. *ONR tech. Rep.*, 1957, No. 19. (Contract N8-onr-66216, Univer. Minnesota.)

OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 299–301.

OSGOOD, C. E. An investigation into the causes of retroactive interference. *J. exp. Psychol.*, 1948, **38**, 132–154.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.

RUSSELL, W. A., & JENKINS, J. J. Studies on the role of language in behavior: The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *ONR tech. Rep.*, 1954, No. 11. (Contract N8-onr-66216, Univer. Minnesota.)

RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

RYAN, J. J. An experimental comparison of response transfer facilitated by meaningfully similar and associated stimuli. Unpublished doctoral dissertation, Univer. Minnesota, 1957.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers College, 1944.

UNDERWOOD, B. J. Associative transfer in verbal learning as a function of response similarity and degree of first list learning. *J. exp. Psychol.*, 1951, **42**, 44–53.

(Received November 20, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 6, 416–418

SUPPLEMENTARY REPORT: ALTERATION IN THE REINFORCEMENT VALUE OF A POSITIVE REINFORCER

ALEXANDER M. BUCHWALD[1]

Indiana University

Data from several experiments (Buchwald, 1959a, 1959b) indicate that an event which consists of *E* saying nothing (N) after *S*'s response can acquire reinforcement value when N is an alternative to "Right" (R) or to "Wrong" (W) in a reinforcement combination. It is possible that this effect may be confined to the event, N, or to events which have zero reinforcement value at the start of an experimental session. The present paper reports an attempt to alter the reinforcement value of an event, *E* saying "mmhm" (H), which has been demonstrated to have a positive reinforcement value for the subject population.

Method.—In Phase I, "induction training," each *S* learned under one of the three different reinforcement combinations: RH, HW, or RW. In Phase II, "test training," all *S*s learned a second task under the same reinforcement combination, namely, "mmhm"-nothing (HN). Twenty *S*s from the introductory psychology course were assigned to each of the three induction training groups. The first 144 cards of the decks used in the previous experiments were used in Phase I. Each *S* was run (individually) until he had either made correct responses on two consecutive blocks of 18 cards, or until he had responded to the entire deck. Phase II followed immediately, without break in procedure, and a second deck of 48 cards was used to give 48 trials. These cards differed from the Phase I deck only in that new nonsense syllables were used and only one number was combined with a given syllable pair. Instructions and other details of procedure were identical with those of the earlier experiments.

Results and discussion.—Both mean and median numbers of correct responses made during test training placed the groups in the order HW, RW, RH (from least to most). Use of the Kruskal-Wallis and Mann-Whitney tests indicated that the groups differed significantly at the 5% level except for the RW vs. RH comparison. These effects cannot be attributed solely to alterations in the reinforcement value of (H) since the response measure is markedly influenced by errors which occur prior to *S* making his first correct response on a given pair. Such errors, since they precede the first reinforcement of a correct response by "mmhm," do not depend on the reinforcement value of "mmhm" but must be due to other aspects of the experimental situation.

A purer picture of the reinforcement values of "mmhm" is given by an analysis in which all trials for each *S* preceding the occurrence of the first "mmhm" on a given syllable pair are discarded. The ratio of incorrect responses to trials, on the remaining trials, was averaged for each *S*. (This latter step, which yields one measure for each *S*, was necessitated by the fact that some *S*s gave no correct responses on one or more pairs.) Summary statistics are shown in Table 1. The median test on these data revealed that the groups differed, overall, at the .01 level of significance, and that Group RH differed from each of the other groups at the .01 level.

Two pieces of evidence suggested that for *S*s in Group RH the reinforcement value of "mmhm" did not merely decline toward zero, but actually became negative. First, 19 of 20 *S*s in Group RH, as compared to only 19 of 40 *S*s in the other two groups, made at least one incorrect response subsequent to the first correct response on a given pair. Further, *S*s in Group RH typically had such a pattern on all syllable pairs while few of the other *S*s did so. Second, many *S*s in Group RH made more incorrect responses as Phase II continued (see Table 2). The χ^2 test applied to Table 2 yielded a value significant at .001. These data suggest that many *S*s in Group RH learned the responses followed by N rather than those followed by H. This finding is all the more striking

TABLE 1

RATIO $\left(\frac{\text{ERRORS}}{\text{TRIALS}}\right)$ FOR TRIALS AFTER FIRST CORRECT RESPONSE ON A GIVEN PAIR

Group	Mean	Mdn.	*SD*
RH	.509	.535	.343
HW	.059	.025	.070
RW	.107	.000	.108

[1] The author is grateful to David Phillips who collected the data for this experiment.

TABLE 2

NUMBER OF *Ss* SHOWING AN INCREASE, A DECREASE, OR NO CHANGE IN NUMBER OF CORRECT RESPONSES FROM TRIALS 1–12 TO TRIALS 37–48

Group	Number of *Ss*		
	Increase	Decrease	No Change
RH	8	10	2
HW	15	1	4
RW	11	1	8

when one considers that N should be more likely to have a positive reinforcement value for those *Ss* who heard *E* say "Wrong" in the experimental situation, i.e., *Ss* in Groups RW or HW, than for the *Ss* in Group RH.

The results indicate that the induction of reinforcement value through contrast with some alternative reinforcing event can occur for events other than nothing, and is not confined to events which have zero reinforcement value at the start of the experimental session.

REFERENCES

BUCHWALD, A. M. Extinction after acquisition under different verbal reinforcement combinations. *J. exp. Psychol.*, 1959, **57**, 43–48. (a)

BUCHWALD, A. M. Experimental alterations in the effectiveness of verbal reinforcement combinations. *J. exp. Psychol.*, 1959, **57**, 351–361. (b)

(Received May 12, 1960)

REPLICATION REPORT: VERBAL CONTEXT AND THE RECALL OF MEANINGFUL MATERIAL

PATRICIA RICHARDSON AND JAMES F. VOSS

College of Wooster

This report presents the replication of a study by Miller and Selfridge (1950) which indicated that the number of words recalled from the single reading of a list of words varied as a function of both the statistical approximation of the words to English and the number of words in the list. The present experiment also provided for the evaluation of two factors that were not controlled in the the original study.

Method.—A randomly permuted 7×7 latin square design was employed where the letters represented the following approximations to English: 0, 1, 2, 3, 5, 7, and text. Three identical latin squares were employed for 10, 30, or 50 words in the list, respectively. Analysis of the latin square sources of variation provided for evaluation of the ordinal position and sequences of the seven types of verbal material. These two factors were confounded with verbal material in the original study.

Twenty-one groups of eight *Ss* each served with each group constituting a row in one of the three squares. The verbal material was that published by Miller and Selfridge (1950, pp. 184–185). The words for all groups were read by one *E* at a 2-sec. rate. The *Ss* were instructed to listen to the reading of the words, and upon completion of the single reading of a given list, they were instructed to write as many of the words of the list as they could, maintaining the order of the words as closely as possible.

Results and discussion.—Figure 1 presents the results of the present study and the comparable results of the original study. Analysis of variance on the word recall frequency data indicated a significant source of variation in Number of Words ($F = 260.5$, $df = 2/147$; error $MS = 60.0$) and in Order of Approximation ($F = 332.9$, $df = 6/882$; error $MS = 11.2$). These findings indicate a greater recall for the longer lists and higher order approximations, respectively. It should be noted that the variability of the length of list groups increased with increasing length.

The interaction of these variables was also significant ($F = 65.7$, $df = 12/882$; error $MS = 11.2$). The significant interaction indicates that as the number of words in the list increases, a greater advantage in words recalled occurs for the higher-order approximations to English. However, as Fig. 1 indicates, the proportion of words recalled decreases as length of list increases.

The results support the findings of Miller and Selfridge, and in addition indicate that

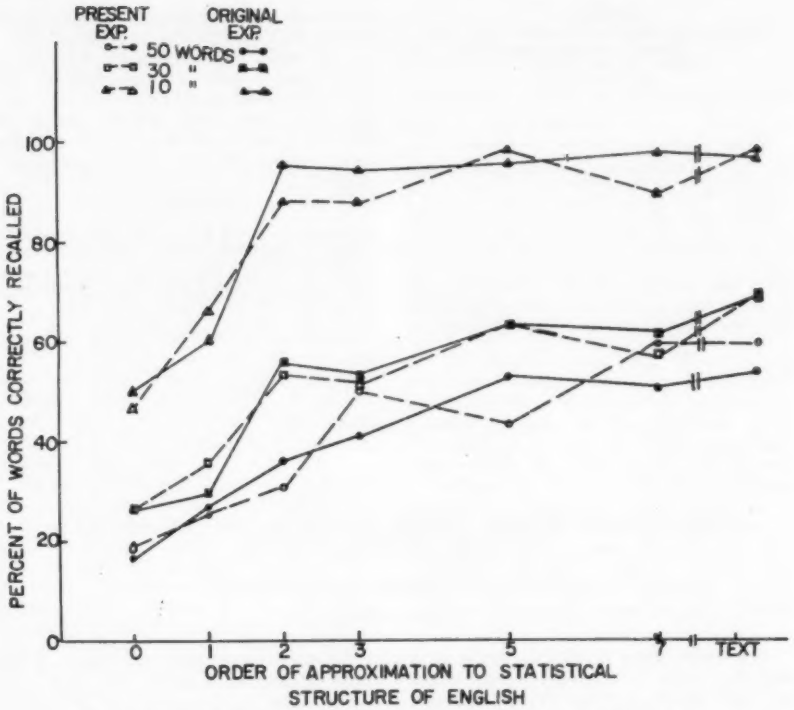

FIG. 1. Percentage of words correctly recalled as a function of order of approximation to English for three lengths of word lists.

the variables of ordinal position and sequence, confounded with verbal material in the original study, do not yield significant differences in words recalled.

REFERENCE

MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.

(Received February 25, 1960)

ERRATA

The article "Supplementary Report: Familiarity and Frequency" by Clyde E. Noble, in the June 1960 issue of this Journal (*J. exp. Psychol.*, 1960, **59**, 432–433), contains an error. On p. 432, tenth line from bottom, second column, the ratio $1/b$ should be $1/a$. This was a typing error and did not affect the computations of the familiarity asymptotes.

In the article "Intersensory Transfer in Learning Sequences" by A. K. P. Sinha and S. N. Sinha (*J. exp. Psychol.*, 1960, **60**, 180–182) the mean trials for the learning the Initial Task in the Visual modality should be 7.525, rather than the 47.525 as shown.

North American
Medical, Health
& Welfare
Telephone Directory

from
Atlantic
to
Pacific
and
Panama
to
Arctic

names
Addresses
Telephone
Numbers
Officials
Suppliers
Institutional
Organizations
Organized Services
Individual Practitioners
etc etc etc

Pre-publication cost
Fifty Dollars with
your order

C. FERGUSON
PO Box 173
Calgary Alberta
Canada

AMERICA'S PSYCHOLOGISTS

A Survey of a Growing Profession

KENNETH E. CLARK

Associate Professor of Psychology
University of Minnesota

A report on a three-year study of American Psychologists, Sponsored by the Policy and Planning Board of the American Psychological Association and the National Science Foundation. The volume presents a view of a growing profession in the mid-twentieth century by describing the people active in psychology. $1.00

Order from

American Psychological Association, Inc.

1333 Sixteenth Street, N.W.
Washington 6, D. C.